1-1-2017

Mechanistic And Inhibitory Analysis Of Clostridium Difficile Associated Diseases

Brianna Marie Jackman
Wayne State University,

MECHANISTIC AND INHIBITORY ANALYSIS OF *CLOSTRIDIUM DIFFICILE* ASSOCIATED DISEASES

by

BRIANNA M. JACKMAN

THESIS

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

DECEMBER 2016

MAJOR: CHEMISTRY (biochemistry)

Approved by:

_____

Advisor Date

DEDICATION

I dedicate this publication to my family

who has supported me, without question,

through thick and thin.

ACKNOWLEDGEMENTS

I would like to acknowledge four mentors who I feel are responsible for molding and guiding me through this academic journey: my high school chemistry teacher, Mr. Jeff Gartrell, who was the first person to instill within me a passion for chemistry; Dr. David Leonard who believed in me when I didn't believe in myself and inspired me to pursue biochemistry in graduate school; Dr. Tamara Hendrickson who has always been there, without judgment, to help me through the hard times; and my advisor, Dr. Andrew Feig, who made this publication possible and has implanted within me the critical thinking skills and confidence to become the scientist that I am today. I thank all of you for your tremendous and unveiling knowledge, support, and guidance. Moreover, I am grateful to all of my lab members throughout the years, especially Adam Boyden and Amit Kumar, who have supported, challenged, and shaped me not only as a scientist, but also as a person. Furthermore, I extend immense appreciation to my friends who have also supported me, but more significantly, taught me the importance of taking time to have fun and enjoy life. Lastly, I would like to express my greatest gratitude to my family: your love and support is the foundation of all I have accomplished and everything that is yet to come.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter One

Chapter Two

Chapter Three

Chapter Four

LIST OF ABBREVIATIONS

A_{340}: absorbance at 340 nm

AB3: Difco™ Antibiotic Medium 3

ACN: acetonitrile

amp: ampicillin

AmpR: ampicillin resistance gene

Anti/Anti: antibiotic/antimycotic

ATP: adenosine triphosphate

***B. meg*:** *Bacillus megaterium*

BL2: Biosafety Level 2

BL21 DE3: strain of chemically competent *E. coli*

***C. diff*:** *Clostridium difficile*

C-18: *n*-octadecane

cam: chloramphenicol

CDAD: *Clostridium difficile*-associated diseases

Cdc42: cell division control protein 42 homolog

CIP: calf intestinal alkaline phosphatase

CO_2: carbon dioxide

CPD: cysteine protease domain

CROP-RBD: C-terminal repetitive oligopeptide receptor binding domain

CXCL1: growth-related oncogene alpha or chemokine (C-X-C motif) ligand 1

D286: aspartate 286

DNA: deoxyribonucleic acid

dNTPs: deoxyribose nucleoside triphosphates

DTT: dithiothreitol

E: enzyme

***E. coli*:** *Escherichia coli*

EDTA: ethylenediaminetetraacetic acid

EI: reversible enzyme/inhibitor complex

EI*: irreversible enzyme/inhibitor complex

ELISA: enzyme-linked immunosorbent assay

EMEM: Eagle's Minimum Essential Media

E_p:E*: final elutions pooled

eq.: equation

FBS: fetal bovine serum

FDA: Federal Drug Administration

FPLC: fast protein liquid chromatography

GH: glucosylhydrolase

GI: gastrointestinal

GT: glucosyltransferase

GTD: glucosyltransferase domain

GTPase: guanosine triphosphate hydrolase enzyme

HEPES-K: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid with potassium counterion

HIV: human immunodeficiency virus

HQS-Epoxy: inhibitor peptide with sequence NH_2-HQSPG$_{epoxy}$HHGGGC$_{amide}$-$CONH_2$

HQS-Parent: inhibitor peptide with sequence NH_2-HQSPG$_{allyl}$HHGGGC-$CONH_2$

HQS-Protected: inhibitor peptide sequence NH_2-HQSPG$_{allyl}$HHGGGC$_{amide}$-$CONH_2$

HQSPWHH: original inhibitor peptide identified by phage display

hTHP-1: human THP-1 monocytes

I: inhibitor

IBD: irritable bowel syndrome

IC_{50}: half-maximal inhibitory concentration

ICE: interleukin-1 converting enzyme or caspase 1

IL-1B: interleukin 1B

IL-6: interleukin 6

IL-23: interleukin 23

IP_6: inositol hexakisphosphate

IPTG: isopropyl β-D-1 thiogalactopyranoside

k_{-1}: rate of dissociation of EI complex

k_{+1}: rate of association of EI complex

k_2: rate constant of EI* formation

kb: kilobase

KCl: potassium chloride

kDa: kilodalton

k_{GH}: glucosylhydrolase enzymatic rate

k_{GH+GT}: combined rate of glucosylhydrolase and glucosyltransferase enzymatic activities

k_{GT}: glucosyltransferase enzymatic rate

K_I: inhibition constant or inhibitor binding affinity constant of EI complex

K_M: Michaelis-Menten constant

LB: lysogeny broth

LDH: lactate dehydrogenase

LPS: lipopolysaccharides

mBMDCs: murine bone marrow derived dendritic cells

mCPBA: meta-chloroperoxybenzoic acid

$MgCl_2$: magnesium chloride

mL: milliliter

mM: micromolar

MRSA: methicillin-resistant *Staphylococcus aureus*

MWCO: molecular weight cut-off

PAMPs: pathogen-associated molecular patterns

pANK: plasmid by Amy N. Kerzmann

pBMJ: plasmid by Brianna M. Jackman

PBS: phosphate-buffered saline

pg: picogram

pNLA: plasmid by Natalie L. Anderson

$(NH_4)_2CO_3$: ammonium carbonate

N_2: liquid nitrogen

Na_2HPO_4: sodium phosphate

NaCl: sodium chloride

NADH: nicotinamide adenine dinucleotide

NADPH: nicotinamide adenine dinucleotide phosphate

NaOH: sodium hydroxide

NEB: New England Biolabs

NF-κB: nuclear factor kappa-light-chain-enhancer of activated B cells

NH_4OH: ammonium hydroxide

N_i: quantity remaining after time t

NK Cells: natural killer cells

nm: nanometer

nM: nanomolar

N_o: initial quantity

OD_{600}: optical density at 600 nm

PaLoc: *C. diff* pathogenicity locus

PCR: polymerase chain reaction

PDB: Protein Data Bank

PEG-6000: polyethylene glycol 6000

PEP: phosphoenolpyruvate

***pfu*:** *Pyrococcus furiosus* DNA polymerase

PK: pyruvate kinase

PxylA': xylose isomerase gene promoter

Rac1: Ras-related C3 botulinum toxin substrate 1

RhoA: Ras homolog gene family member A

RhoA-Glc: glucosylated (inactive) RhoA

RLU: relative light units

RP-HPLC: reverse phase high performance liquid chromatography

rpm: rotations per minute

RT: room temperature

rTcdA^{540}: toxin A truncate containing amino acids 1-540

rTcdA-WT: wild type recombinant *C. diff* toxin A

rTcdA-NVN: recombinant, inactive mutant of *C. diff* toxin A

rTcdB-WT: wild type recombinant *C. diff* toxin B

rTcdB-NVN: recombinant, inactive mutant of *C. diff* toxin B

RXN: reaction

S: RhoA substrate concentration

SDM: site-directed mutagenesis

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEAP: secreted embryonic alkaline phosphatase

SF-EMEM: serum-free Eagle's Minimum Essential Media

t: specified time point

$t_{1/2}$: half-life or time at 50% decay of product

***taq*:** *Thermus aquaticus* DNA polymerase

***tcda*:** gene encoding *C. diff* enterotoxin A

TcdA: *C. diff* enterotoxin A

***tcdb*:** gene encoding *C. diff* enterotoxin B

TcdB: *C. diff* enterotoxin B

***tcdc*:** gene encoding *C. diff* toxin C

***tcde*:** gene encoding *C. diff* toxin E

***tcdr*:** gene encoding *C. diff* toxin R

TD: translocation domain

tet: tetracycline

TetR: tetracycline resistance gene

TFA: trifluoroacetic acid

TLR: toll-like receptor

TNFα: tumor necrosis factor alpha

Top10: strain of electrocompetent *E. coli*

TSS: transcription start site

U: units

UDP-Glc: uridine diphosphate glucose

ug: microgram

uL: microliter

um: micrometer

uM: micromolar

umol: micromoles

UV-VIS: ultraviolet-visible

V: volts

V_i: rate in presence of inhibitor

V_o: rate in absence of inhibitor

v/v: volume per volume ratio

w/v: weight per volume ratio

xylA': xylose isomerase gene

xylR: transcriptional regulatory repressor of xylose isomerase gene

CHAPTER ONE:

Introduction to the *Clostridium difficile* Pathogen and Associated Diseases

Clostridium difficile (*C. diff*, **Figure 1A**) is a Gram-positive, anaerobic pathogen known to form spores and give rise to hospital-acquired *C. diff*-associated diseases (CDAD).[1] Often, the infection is acquired by patient-to-patient transfer and subsequent ingestion of *C. diff* spores known to survive in the hospital setting.[2, 3] CDAD is characterized by a range of symptoms such as excessive diarrhea, colonic inflammation, septic shock, pseudomembranous colitis (**Figure 1C**) and in extreme cases, death.[1, 7] The cycle of CDAD manifestation and recurrence is outlined in **Figure 2**. Essentially, patient treatment with broad-spectrum antibiotics or chemotherapeutic agents[8] disturbs the normal bacterial community within the colon, allowing *C. diff* to colonize the gut and cause infection.[7] Moreover, several antibiotic-resistant strains are continually emerging, only some of which have been characterized.[9, 10] These resistant, hypervirulent strains, as well as the common production of environmentally-tolerant spores[1, 3, 10] escalate relapse rates in previously treated patients.[8] As a consequence of bacterial evolution,

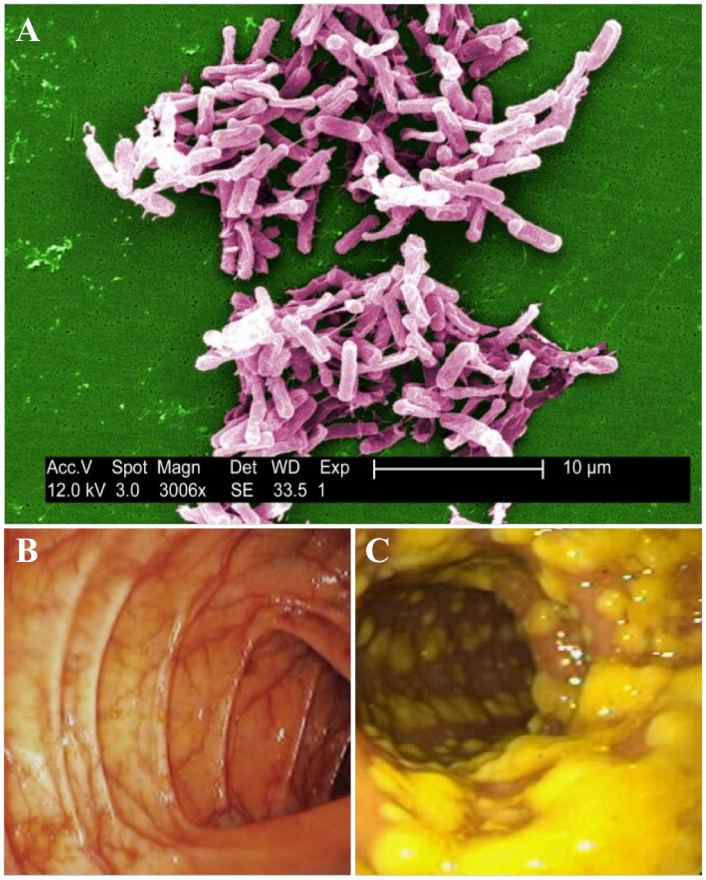


Figure 1: *C. diff* pathogen and pseudomembranous colitis. A: Scaled image of the *Clostridium difficile* microbe.[4] **B:** Image of a normal, healthy colon.[5] **C:** Yellow plaques characteristic of pseudomembranous colitis in a diseased patient.[6]

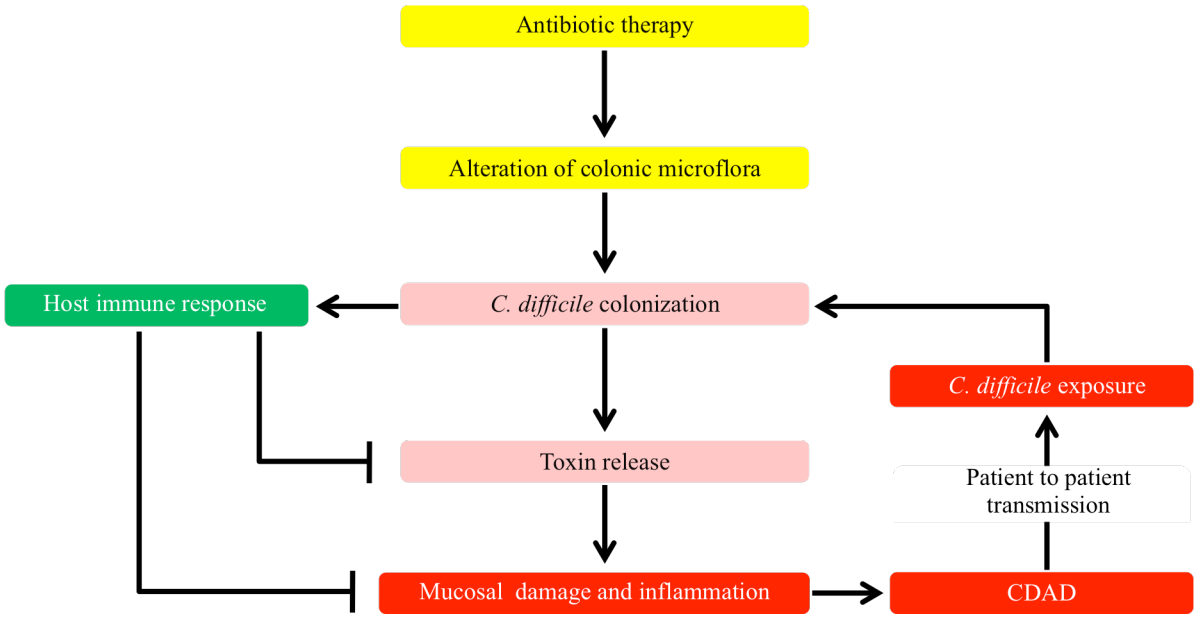


Figure 2: Cycle of CDAD manifestation and recurrence. Broad-spectrum antibiotics lead to the disruption and modification of the normal microbiota in the gut, which subsequently allows *C. diff* to colonize. If the host immune system is efficient enough to fight off the infection, then CDAD will be nullified. However, if the immune system is compromised, *C. diff* will be allowed to release toxins, cause mucosal damage and inflammation within the gut, and eventually evolve fully to CDAD. At this stage, the disease cycle is continued via patient-to-patient transmission.[11]

prevalence and mortality rates associated with CDAD are actively increasing, with susceptibility relying not only on antibiotic consumption and hospitalization, but elderly individuals and those with inflammatory bowel disease (IBD) or human immunodeficiency virus (HIV) are also especially vulnerable.[12] Roughly 500,000 Americans per year acquire CDAD, where 1 out of every 5 patients experience a recurrent infection, and nearly 30,000 of the overall infected patients die within only one month of disease onset.[13] Notably, 80% of yearly deaths due to *C. diff* infection occurred among those aged 65 years and above.[13] Furthermore, the number of CDAD victims has recently outgrown those infected with methicillin-resistant *Staphylococcus aureus* (MRSA), causing CDAD-related medical expenses to reach $4.8 billion in the United States per year.[13, 14]

Enterotoxins A (TcdA) and B (TcdB), shown in **Figure 3**, are the main virulence factors produced by *C. diff*.[10] Both TcdA/B are composed of four domains: an enzymatic glucosyltransferase domain (GTD), a cysteine protease domain (CPD), a central translocation domain (TD), and a receptor-binding region at the c-terminus, also known as the C-terminal repetitive oligopeptide receptor-binding domain (CROP-RBD).[15] The *tcda* and *tcdb* genes encoding Toxin A and B, respectively, are located within a 19.6 kb chromosomal region known

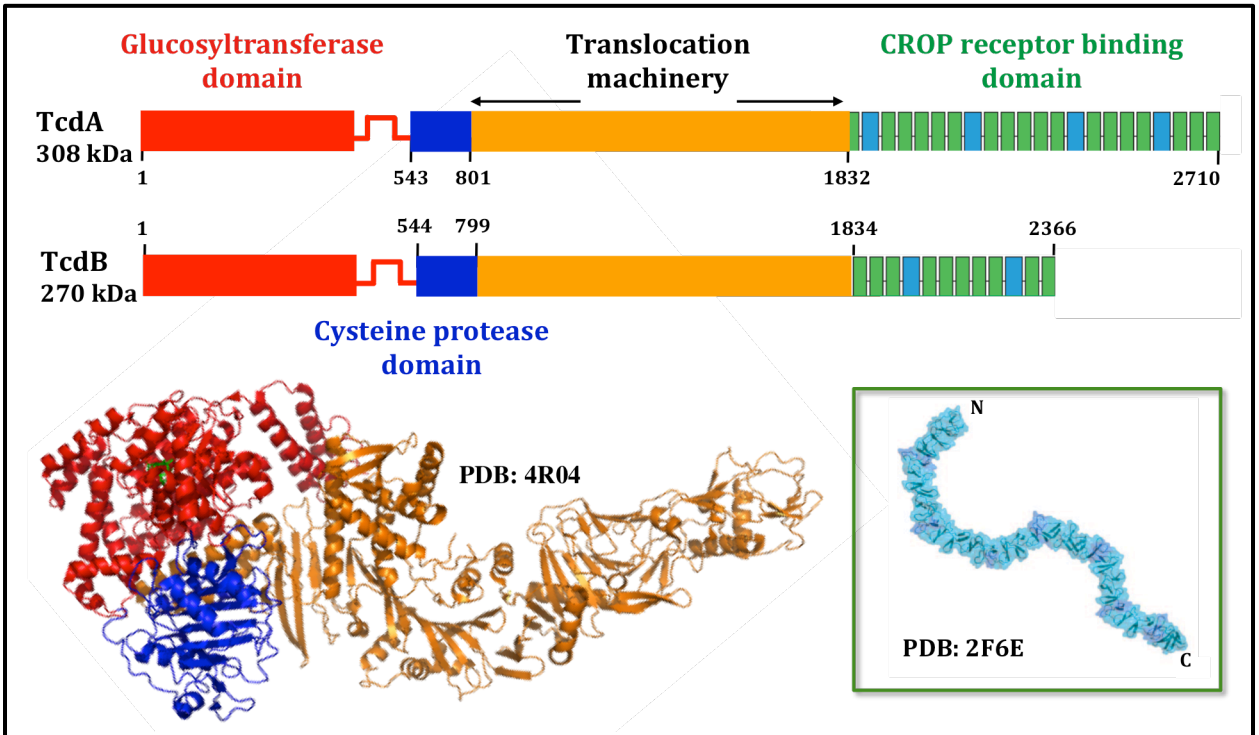


Figure 3: *Clostridium difficile's* primary virulence factors. The proteinaceous enterotoxins A (TcdA, top) and B (TcdB, bottom) produced by *C. diff* are shown above with corresponding molecular weights to the left. Both toxins are composed of four domains (GTD, red; CPD, blue; TD, orange; CROP-RBD, green) and differ mostly in the length of their CROP-RBD regions. Each domain is labeled with the chronological number of amino acid residues, beginning at the N-terminus. The recently determined crystal structure of the first three domains is shown below (PDB: 4R04)[16] as well as a segment of the TcdA CROP-RBD (PDB: 2F6E), which consists of 32 short repeats and 7 interspersed long repeats (light blue).[17, 18]

as the pathogenicity locus (PaLoc).[19] Also present within the PaLoc is *tcdr* and *tcdc*, which are genes that code for a positive and a negative transcriptional regulator of toxin production, respectively, and *tcde*, a holin-like protein necessary for secretion of the toxins into the

surrounding environment. These three additional PaLoc genes, although extremely important, are only a few of many factors involved in the regulatory networks of TcdA/B synthesis.[19]

Pathogenesis of TcdA/B is exceptionally fascinating; these holotoxins possess all necessary machinery required to infect and kill cells. The mechanism of action (**Figure 4**), although yet to be completely understood, proceeds by first utilizing the CROP-RBD to bind a glycan receptor present at the surface of colonic epithelium cells.[20] More specifically, the CROP

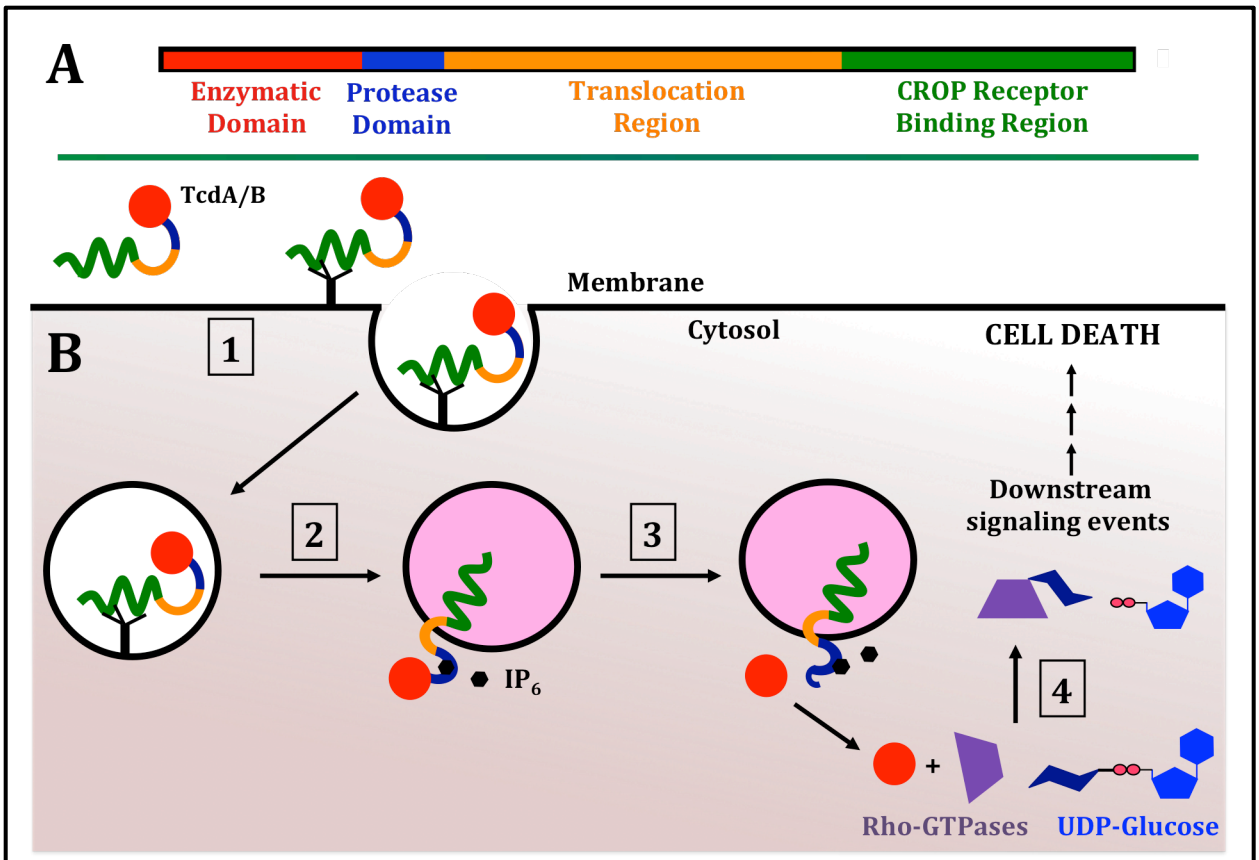


Figure 4: Mechanism of TcdA/B pathogenesis. A: Diagram of TcdA domains. **B:** Pathogenesis begins when the CROP-RBD (green) binds a receptor and initiates endocytosis (1). As the endosomal pH drops, the TD (orange) forms a pore-like structure that traverses the endosomal membrane. Next, the N-terminus translocates into the cytosol (2), IP_6 (black hexagons) causes self-cleavage of the CPD (blue) (3), and the GTD (red) is released. Finally, the GTD enzymatically glucosylates Rho-GTPase proteins (purple) and causes cell death (4).[21]

has portrayed specific binding affinity to glycan moieties ending in a Galα1,3Galβ1,4GlcNAc sequence, where the α1,3 linkage amidst the two galactose sugars is suggested to be of importance.[22] Upon binding, receptor-mediated endocytosis is activated and results in an

endosome-encapsulated toxin.[23] The acidic environment, which is characteristic during endosomal maturation, then allows the TD to unfold, expose its hydrophobic residues, and insert into the endosomal membrane, forming a pore-like structure that acts as the pathway for endosomal release.[20, 24] Beginning at the N-terminus, the GTD and CPD are fed through the pore and into the cytosol.[24, 25] The mechanism of endosomal insertion and release is still not completely known, but is believed to proceed in a way very similar to that of the anthrax toxin produced by *Bacillus anthracis*.[26] Upon entering the cytosol, the CPD binds native inositol hexakisphosphate (IP_6), which causes a conformational change allowing self-cleavage of the CPD and release of the enzymatic GTD.[24] The GTD is then free to exploit native UDP-glucose (UDP-Glc) as a substrate by hydrolyzing UPD-Glc and transferring the glucose moiety to threonine 37 of Rho GTPase family members (RhoA, Rac1, Cdc42) residing within intestinal epithelium cells.[27] Glycosylation of these small proteins (5-20 kDa) inactivates the enzymes catalytic GTP-ase behavior, resulting in loss of tight junctions between cells, loss of intestinal barrier function, cell death, and the characteristic symptoms of CDAD.[8]

The vigorous manifestation of *C. diff* and production of toxins also causes activation of the host immune system.[28, 29] As a first line of defense, the innate immune system generates non-specific immunity by producing small proteins, called cytokines, which lead to numerous downstream signaling events within the cell.[30] Cytokines, such as interleukins and chemokines, work together to identify and remove foreign matter by recruiting immune cells to the infection site, balancing humoral and cell-based immunity, and activating the adaptive immune system.[30, 31] A compromised immune system in combination with existing *C. diff* infection is unfavorable and will likely progress illness within the patient, however an overactive immune system can also be destructive by initiating an excessive inflammatory response that may damage healthy

tissue. Therefore, although innate immunity is designed to fight off infection, when regarding *C. diff* and CDAD patients, in fact the system most often produces a detrimental amount of colonic inflammation that enhances disease progression.[2, 32] Consequently, the problem and/or cure of CDAD becomes two-fold; both host and pathogen play a synergistic role in *C. diff* disease pathogenesis and progression.[32] In other words, inhibition of bacterial colonization and toxic virulence factors, as well as down-regulation of harmful inflammation in response to *C. diff* bacterial components is needed to elicit an efficient therapeutic remedy.

A few of the current treatment methods to combat CDAD include antibiotic therapy, fecal transplantation, probiotics, and immunotherapy.[1, 33, 34] Antibiotics, such as vancomycin and metronidazole, are usually the first course of action against the *C. diff* pathogen, however this approach has shown to have high recurrence rates and is also responsible for numerous resistant and hypervirulent strains.[1] Probiotics and fecal transplantation are both methods that attempt to reestablish the normal microflora within the gut, which is essential for diminishing *C. diff* and curing the patient. Although probiotic and fecal transplantation therapies have had some success, they both have downfalls as well. Probiotics, which are still considered to be controversial (not always FDA regulated), are not typically potent enough to cure the infection on their own and therefore are usually administered in conjunction with other treatments.[33, 34] On the other hand, fecal transplantation has had many good outcomes, however it is known to have low patient compliance, moderate recurrence rates, and risk of transferring disease or other physiological traits (obesity, depression, etc.) to the patient.[1, 34, 35, 36]

In attempt to sequester and fight off toxins, immunoglobulin and/or monoclonal antibodies have had variable outcomes, however these therapies are often administered at extremely high, sometimes toxic doses and again, usually in conjunction with other treatment

methods.[1, 33, 34] It would be ideal to avert the disease altogether, however there has yet to be a preventative *C. diff* vaccine to pass clinical trials.[1] With all things considered, the current treatment options are ineffective and neglect to simultaneously target the bacterial pathogen and the inflammation-causing virulence factors that have already been produced. Moving forward, an effective therapeutic method that will inhibit all factors of CDAD is needed.

The specific goals of my thesis investigate GTD inhibition, host and pathogen contribution to immune system activation, and production of active rTcdB-WT. To begin, the first project deals with the development of an epoxide-containing peptide (HQS-Epoxy) as a novel, covalent inhibitor of the glucosyltransferase activity of *C. diff* toxins A and B (**Chapter 2**). Here, I aim to elucidate the rate of relative inhibition and crosslinking of the epoxide inhibitor to the GTD, the half-maximal inhibitory concentration (IC_{50}) of the peptide, and visual confirmation of cell protection against rTcdA-WT when inhibitor is present. Secondly, in regards to the host immune system, a collaboration was formed with Dr. William Petri's lab at the University of Virginia to elucidate the relationship between *C. diff* bacteria/toxin exposure and inflammasome activation within a murine host. **Chapter 3** details *in vitro* and *in vivo* activation of inflammation and cytokine production, as well as *in vivo* cellular damage, in response to active rTcdA-WT. Finally, since toxin B was not yet available in the Feig Lab, **Chapter 4** outlines the cloning, expression, and purification of rTcdB-WT and rTcdB-NVN (catalytically inactive mutant), as well as *in cellulo* confirmation of rTcdB-WT cytotoxicity. Together, these studies can facilitate characterization of CDAD as a whole, as well as distinguish the potential of our peptide inhibitors, and give further insight to prospective therapeutic methods.

CHAPTER TWO:

Irreversible Inhibition of *Clostridium difficile* Toxin A Glucosyltransferase Activity via an Epoxide-Containing Peptide

As described in **Chapter 1**, both *C. diff* and the virulence factors TcdA/B are responsible for activating the host innate immune system, initiating a detrimental inflammatory response, and causing excessive damage to the colon.[37, 38] Current treatment methods, most of which are antibiotics that target *C. diff*, are only modestly effective at curing CDAD. As a result, methods of targeting and inhibiting TcdA/B directly have become a topic of interest. Each of the four domains of TcdA/B (GTD, CPD, TD, and CROP-RBD) play separate and significant roles within the mechanism of infection, thus any of them can be targeted for inhibition.

With hopes of impeding initial CROP receptor-binding and entry of TcdA/B into host cells, Cherian, *et al.,* have recently engineered a cell line that stably expresses glycan derivatives which can be used to identify potential therapeutics by determining specific, high-affinity glycan receptor-binders.[22] Studies have also identified specific regions within the TD that are imperative for endosomal release of the toxin[39] as well as molecules that disrupt pore formation and/or translocation across the membrane, either structurally or by inhibiting endosomal acidification, to culminate in an encapsulated toxin that is essentially nontoxic.[40, 41] Furthermore, several molecules have been reported to inhibit autoprocessing/cleavage of the CPD, activate premature cleavage of the CPD (before entering host cell),[42, 43, 44] and inhibit enzymatic activity of the GTD,[41, 45, 46] and although numerous attempts to obstruct TcdA/B on a molecular level have revealed a better understanding of pathogenesis, none have led to a therapeutic antitoxin remedy that is effective towards curing CDAD in humans.

Interested by the inherent and newly discovered inflammatory damage to the colon that is the result of the GTD, the Feig Lab is currently investigating methods of TcdA/B

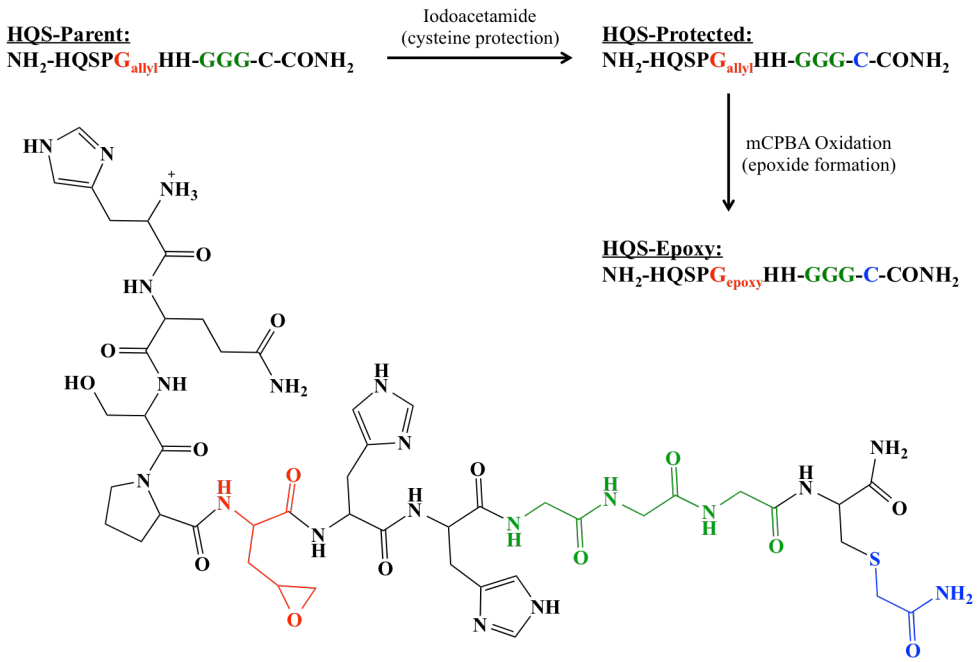


Figure 5: Structure and production pathway of the HQS-Epoxy peptide inhibitor. The parent peptide (HQS-Parent) is purchased with an allylglycine at the fifth position (red), a three-glycine linker (green) for flexibility, and a cysteine residue for future labeling. Iodoacetamide is used to amidate and protect the C-terminal cysteine (blue) before treating with mCPBA to oxidize the allyl group to an epoxide at position five.

glucosyltransferase inhibition. Previously, Dr. Sanofar Abdeen used the phage display technique to identify a heptapeptide (HQSPWHH) which exhibited nanomolar binding affinity to the GTD of *C. diff* rTcdA.[11, 46] Transitioning from phage to free peptide involved addition of a flexible glycine linker and a C-terminal cysteine for future labeling and binding purposes, resulting in the amino acid sequence HQSPWHHGGGC. As expected, this peptide displayed inhibition of glucosyltransferase activity *in vitro*, however further studies unexpectedly showed that it lacked the ability to protect against rTcdA *in cellulo*,[11] which was hypothesized to be the result of inhibitor dissociation during endosomal translocation.

To prevent dissociation and create a covalent connection between the inhibitor and the GTD active site, the tryptophan at position 5 was replaced with allylglycine (HQS-Parent) and subsequently modified to comprise a reactive epoxide moiety (HQS-Epoxy) with cross-linking ability.[11] The structure and derivatization process of HQS-Epoxy is depicted in **Figure 5**. Further

analysis of HQS-Epoxy determined that incorporation of the epoxide did indeed generate an irreversible inhibitor that was confirmed by mass spectrometry to cross-linked to the catalytic region of rTcdA-GTD[11, 47] and after incubation of 30 minutes successfully protected against rTcdA *in cellulo,* demonstrating 95% cell protection at a concentration of 600 μM and a half-maximal inhibitory concentration (IC_{50}) of ~100 μM.[11, 46] Kinetically, the binding of HQS-Epoxy to rTcdA is a two step process that combines reversible and irreversible inhibition as described in **equation 1**,

$$E + I \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} EI \xrightarrow{k_2} EI^* \qquad (\boldsymbol{eq.1})$$

that is partially competitive to UDP-Glc and RhoA substrates, where the inhibitory binding affinity constant (K_I) for the reversible, loosely bound EI complex is defined as k_{-1}/k_{+1} and k_2 is the rate constant (i.e., cross-linking rate or rate of inactivation) corresponding to the tightly bound, covalent EI* complex containing inactivated enzyme. Because of its potential to dissociate (k_{-1}), the initial EI complex formation is assumed to be a slower process than the subsequent time-dependent EI* complex formation which is irreversible and results in a loss of active enzyme over time.

To further analyze these kinetic parameters and extent of glucosyltransferase inhibition generated by HQS-Epoxy, the optical coupled-enzyme assay displayed in **Figure 6A** was employed. With UDP being a common product of both GH and GT enzymatic reactions performed by rTcdA/B, this assay couples pyruvate kinase (PK) and lactate dehydrogenase (LDH) to allow for the simultaneous detection of the individual rates by monitoring the stoichiometric loss of NADH over time at an absorbance of 340 nm. **Figure 6B** depicts that

when RhoA or a suitable GTPase is present, TcdA/B exploits both GH and GT functionality and displays a rate that is the additive effect of both enzymatic reactions (k_{GH+GT}). However, when

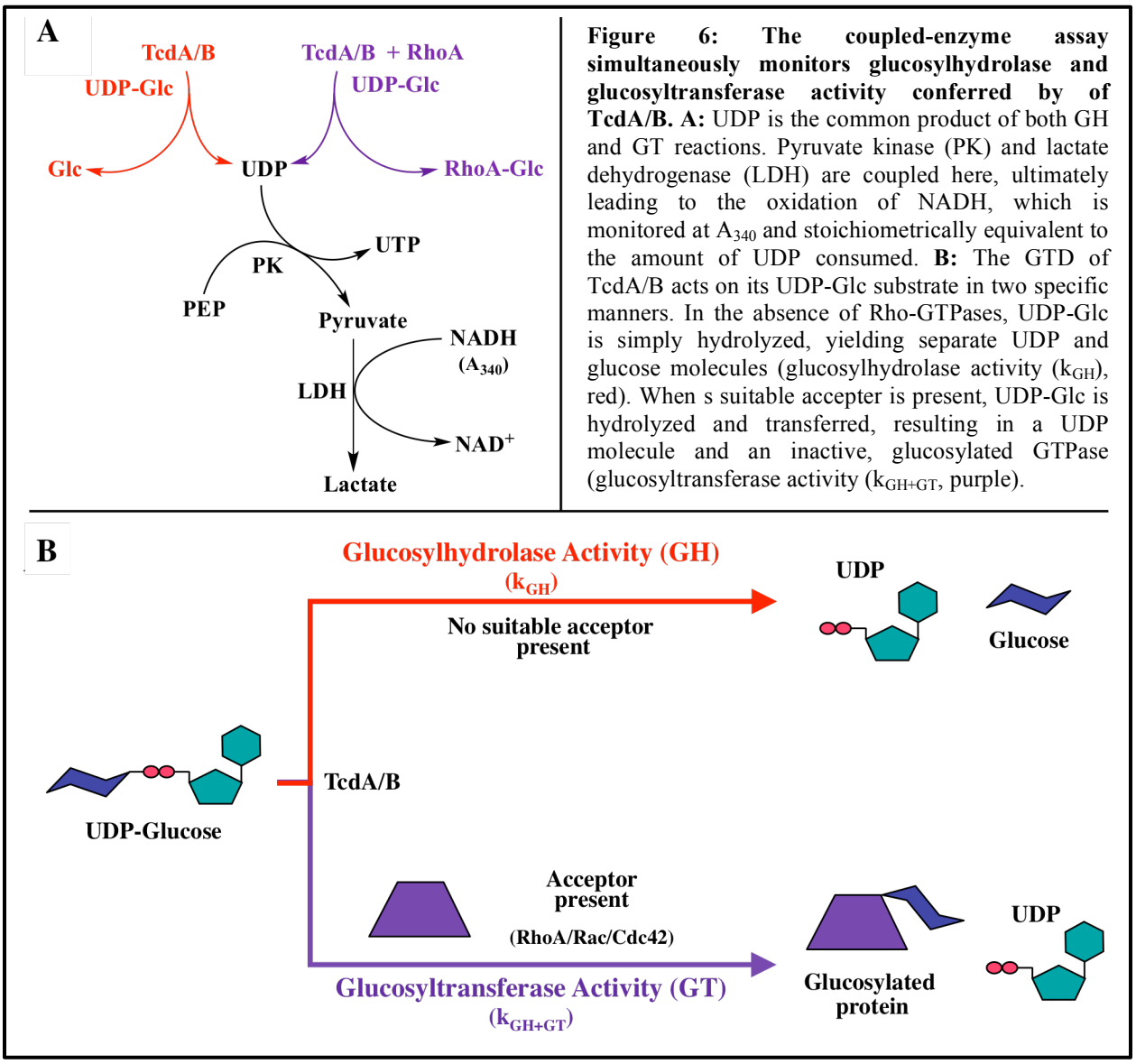


Figure 6: The coupled-enzyme assay simultaneously monitors glucosylhydrolase and glucosyltransferase activity conferred by of TcdA/B. A: UDP is the common product of both GH and GT reactions. Pyruvate kinase (PK) and lactate dehydrogenase (LDH) are coupled here, ultimately leading to the oxidation of NADH, which is monitored at A_{340} and stoichiometrically equivalent to the amount of UDP consumed. **B:** The GTD of TcdA/B acts on its UDP-Glc substrate in two specific manners. In the absence of Rho-GTPases, UDP-Glc is simply hydrolyzed, yielding separate UDP and glucose molecules (glucosylhydrolase activity (k_{GH}), red). When s suitable accepter is present, UDP-Glc is hydrolyzed and transferred, resulting in a UDP molecule and an inactive, glucosylated GTPase (glucosyltransferase activity (k_{GH+GT}, purple).

RhoA is not present or has been diminished, the resulting rate is that of the GH reaction (k_{GH}) alone, which can then be subtracted from k_{GH+GT} to determine k_{GT}. Pre-incubation of toxin with various concentrations of HQS-Epoxy inhibitor, and for specified times prior to assay initiation, was done to allow for a time-dependent analysis of GT inhibition and is discussed further below.

When using rTcdA-WT in the coupled-enzyme assay the slope of k_{GH+GT} was determined to be -6x10^{-5} µmol/min (1.85 turnovers/min), as seen in **Figure 7A/B**, which is hypothesized to be the structural result of the holotoxin TD and CROP folding back and interacting with the GTD

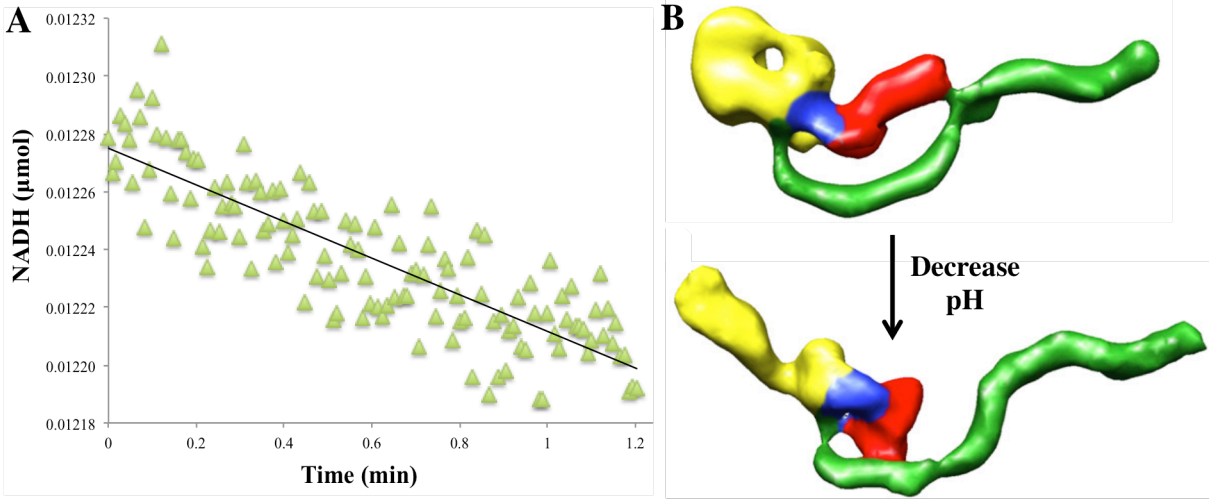


Figure 7: k_{GH+GT} rate of rTcdA-WT as determined by the coupled-enzyme assay and visual occlusion of rTcdA-GTD at high pH. A: Initial NADH turnover rate of rTcdA-WT (1.85 turnovers/min) where the slope of k_{GH+GT} is -6x10^{-5} µmol NADH/min. **B:** Negative stain EM images reconstructed in 3D show structural changes in rTcdA upon decreasing pH.[48, 49] GTD: red. CPD: blue. TD: yellow. CROP-RBD: green.

and CPD in a manner that obstructs enzymatic activity prior to acidification and cleavage.[16, 48, 49] Incubation of rTcdA-WT with IP_6 has shown to induce cleavage and allow rTcdA-WT activity *in vitro*,[48] however this was not attempted for the coupled-enzyme assay to avoid adding another enzymatic rate to the equation. Instead, in order to successfully analyze the GH and GT rates of *C. diff* toxin A, the GTD was isolated as a truncated construct containing the first 540 amino acids of the holotoxin (rTcdA^{540}, **Figure 8A**).[50, 51] As opposed to rTcdA-WT, rTcdA^{540} is fully

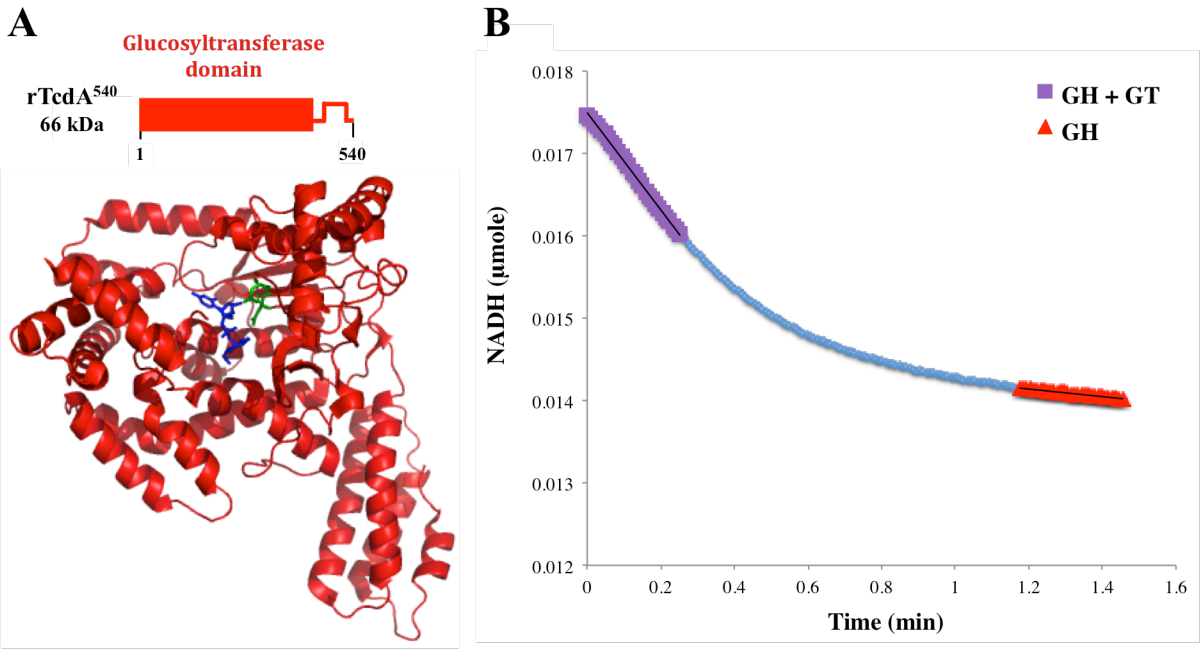


Figure 8: Structure and distinct enzymatic rates of rTcdA^{540}. A: Shown on top, rTcdA^{540} contains amino acids 1-540 of the TcdA holotoxin producing a 66 kDa protein. Below, the structure of the GTD (PDB: 3SRZ)[51] is displayed with catalytic residues D286 and D288 in green and UDP-Glc in blue. **B:** Example of data obtained from the coupled-enzyme assay using rTcdA^{540}, where k_{GH+GT} is shown in purple with a slope of -0.00590 µmol/min (181.5 turnovers/min) and k_{GH} is shown in red with a slope of -0.00046 µmol/min (14.2 turnovers/min).

active when used in the coupled-enzyme assay and gives distinct rates for GH and GT (**Figure 8B**), which have also been previously reported.[11]

Although analysis of rTcdA inhibition via HQS-Epoxy was monitored *in cellulo*,[11,46] the overall extent of inhibition, crosslinking rate (k_2), and additional kinetic parameters utilizing the coupled-enzyme assay and rTcdA^{540} needed further investigation. Since the GH rate is slow and minimally affected by HQS-Epoxy, an average GH rate of 15 ± 1 turnovers/min (n = 4) was determined for rTcdA^{540} in the absence of inhibitor (data not shown). Once observed k_{GH+GT} rates in the absence and presence of varying HQS-Epoxy concentrations were determined, the average GH rate of 15 ± 1 turnovers/min was subtracted to give the average initial GT turnover rates seen in **Figure 9A**. At time zero, the GT rate of rTcdA^{540} (~140 turnovers/min) showed little change regardless of inhibitor concentration, but continuous incubation of rTcdA^{540} with HQS-Epoxy at

37 °C resulted in decreasing GT activity over time, indicating that the crosslinking process is not occurring instantaneously and is in fact time-dependent. **Figure 9A** also shows that although rTcdA^{540} lost roughly 50% activity over the 60 minute time course in absence of inhibitor, there was nearly a 9.5 fold decrease in GT rate when in the presence of 600 µM inhibitor for the same

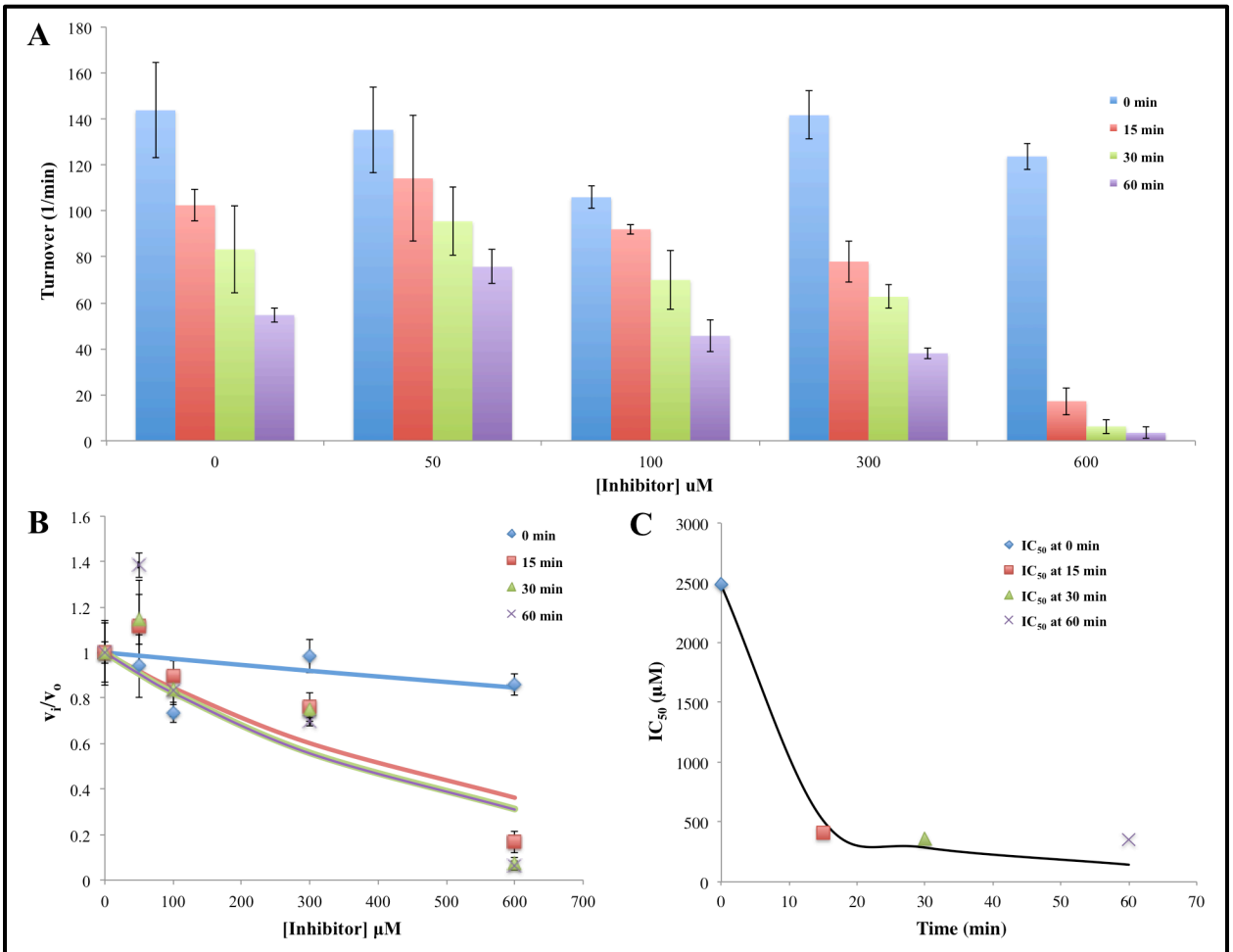


Figure 9: rTcdA^{540} glucosyltransferase rates for varying inhibitor concentrations over time. A: rTcdA^{540} (325 nM) was incubated with 0, 50, 100, 300 or 600 µM inhibitor at 37 °C for the duration of 1 hour. Aliquots from each sample were taken at 0 (blue), 15 (red), 30 (green), and 60 min (purple) and added to the coupled-enzyme assay to determine glucosyltransferase activity, shown here as inverse minutes. **B:** Relative rates (v_i/v_o) for varying HQS-Epoxy concentrations (µM) at 0, 15, 30, and 60 minute time points. **C:** Percent inhibition of 300 µM (green) and 600 µM (purple) over time. All error bars indicate standard deviation; n = 3.

time period. These data suggest that rTcdA^{540} loses enzymatic activity over the 60 minute time course when incubated at 37 °C. However, this effect is not the sole reason for the decrease in GT rate that is seen in presence of the inhibitor.

To account for loss of rTcdA^{540} activity over time, **Figure 9B** portrays the relative rate (v_i/v_o) as a function of the corresponding pre-incubation time where v_i is the GT rate in presence of inhibitor, v_o is the GT rate in absence of inhibitor, and v_i equals v_o when [inhibitor] = 0. Representation of the kinetic data in this manner visually clarifies that 600 μM inhibitor results in roughly 95% inhibition, agreeing with previously reported *in cellulo* data.[11] **Figure 9B** also exhibits that nearly 100% of potential crosslinking occurs in less than 15 minutes as longer incubation times of 30 and 60 minutes did not confer additional inhibition, thus suggesting an upper limit of 1.3 min^{-1} for the rate constant k_2. Supplementary data at time points below 15 minutes are necessary to more accurately define the rate constant of the crosslinking event.

Nevertheless, IC_{50} values for the specified pre-incubation times were obtained by fitting the data in **Figure 9B** to a generic half-life equation,

$$N_t = N_o \cdot \left(\frac{1}{2}\right)^{\frac{t}{t_{1/2}}} \qquad (\boldsymbol{eq.2})$$

where N_t is the quantity remaining after time t, N_o is the initial quantity, and $t_{1/2}$ is the half-life. In this case however, when relative rates are plotted with inhibitor concentration on the x-axis, $t_{1/2}$ then becomes the inhibitor concentration at 50% of the initial quantity, also known as the half-maximal inhibitory concentration or IC_{50}. Following Krippendorff, *et. al.*,[52] to determine K_I and k_2 variables, time-derived IC_{50} values were then graphed against time (**Figure 9C**) and fit by non-linear least squares regression to **equation 3**,

$$IC_{50}(t) = K_I\left(1 + \frac{S}{K_M}\right) \cdot \left(\frac{2 - 2e^{-\eta IC_{50}} \cdot k_2 \cdot t}{\eta IC_{50} \cdot k_2 \cdot t} - 1\right) \quad with \qquad (\boldsymbol{eq.3})$$

$$\eta IC_{50} = \frac{IC_{50}(t)}{K_I\left(1 + \frac{S}{K_M}\right) + IC_{50}(t)}$$

where $IC_{50}(t)$ is the half-maximal inhibitory concentration at time t, S is the substrate concentration (RhoA), and K_M is the Michaelis-Menten constant for RhoA.[11, 50] All variables used and/or derived from fitting the model **equations 2** and **3** can be seen in **Table 1**.

Table 1: ***Kinetic variables for HQS-Epoxy inhibition of rTcdA^{540}.***

Time-derived IC_{50} Values (µM)				**S (µM)[a]**	**K_M (µM)[b]**	**K_I (µM)**	**k_2 (min^{-1})**
0 min	**15 min**	**30 min**	**60 min**				
2487.7	409.4	357.6	355.2	50	300	1066.5	1.3
[a]RhoA substrate concentration used in assays. [b]Michaelis-Menten constant for RhoA affinity to TcdA[11, 50]							

In comparison to previously reported values,[11] the obtained K_I of ~1100 µM is roughly 3.5 orders of magnitude larger than that found for HQS-Parent (K_I = 300 nM) and one order larger than the IC_{50} (100 µM) of HQS-Epoxy.[11] An increase in the K_I value was expected due to decreased binding affinity of the modified peptide following addition of the epoxide moiety, however it is also important to note that these values were obtained using different experimental methods and fitting models. Furthermore, confirming that crosslinking was nearly complete by the 15 minute time point (**Figure 9B**), combined with the lack of GT data corresponding to critical time points below 15 minutes, the k_2 value was somewhat expected to be inaccurate. Presumably, the determined k_2 of 1.3 min^{-1} is seemingly slow, reflecting an upper limit of this value and indicating that EI* is formed at a constant rate of 1.3 crosslinking events per minute. Again, this data is unable to be published because time points below 15 minutes are necessary to accurately determine the values of K_I and k_2.

To visualize if HQS-Epoxy confers cell protection against rTcdA at an incubation time point of zero, toxin and inhibitor were mixed and immediately subjected to Vero cells for 1 hour before imaging. Contrary to the GT kinetics reported above, **Figure 10** shows that even with zero pre-incubation time allotted for crosslinking, the majority of cells were alive and protected from the rounding effects of rTcdA, but in the absence of HQS-Epoxy, cells containing 0.2 nM rTcdA were severely disrupted and nearing death after only 1 hour. These data verify the fact that HQS-Epoxy crosslinks to the GTD in less than 15 minutes and most likely does so following CROP-receptor binding and endosomal acidification, as the TD and CROP interact with the GTD prior to these events,[16, 48, 49] which was corroborated by the lack of activity for rTcdA-WT in the coupled-enzyme assay. Moreover, incubation of Vero cells with 600 μM HQS-Epoxy alone determined that the epoxide-containing peptide inhibitor is in fact not toxic to cells and further validates its potential use as an antitoxin therapeutic.

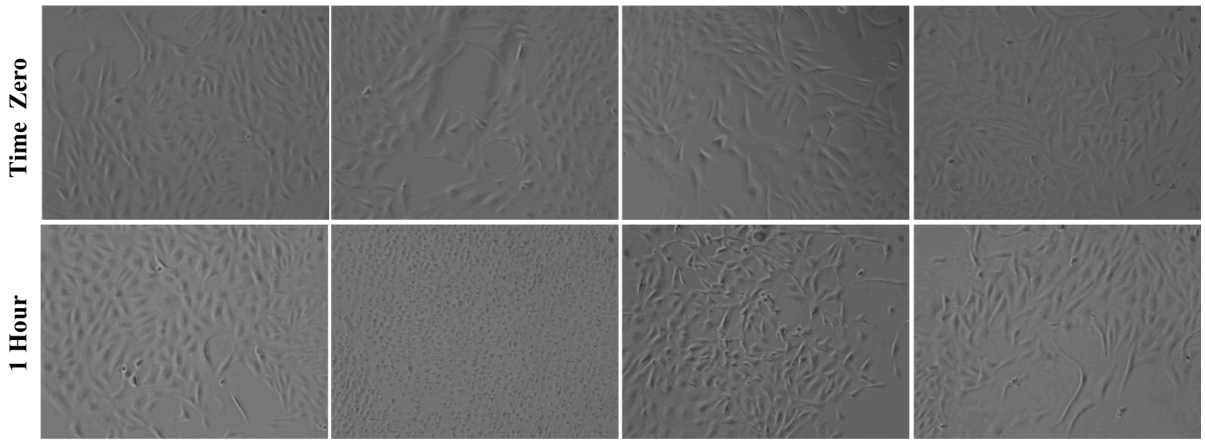


Figure 10: Visual cell protection against rTcdA in presence of HQS-Epoxy inhibitor. Vero cells were grown to confluency at 37 °C and 5% CO_2 and images were taken at time zero (top row) before addition of 0.2 nM rTcdA-WT (column 2), 0.2 nM rTcdA-WT and 600 μM inhibitor (column 3; no preliminary incubation time allotted), or 600 μM inhibitor alone (column 4). Cells were then incubated at 37 °C and 5% CO_2 for 1 hour before imaging again (bottom row). After 1 hour, cells in column 1 (control; media only) show no change as expected. Cells in column 4 remain flat and adherent, indicating that 600 μM inhibitor does not seem to harm cell viability. Following incubation with 0.2 nM rTcdA-WT, Vero cells are rounded, detaching from surface, and showing signs of cell death, however, majority of Vero cells are protected and remain adherent with addition of inhibitor.

The contrast of *in vitro* and *in cellulo* inhibitor effects at time zero are not extremely surprising, however; since the inhibitor peptide cannot bind rTcdA-WT before receptor binding and/or conformational change from endosomal acidification, there is an intrinsic delay in EI and EI* complex formation that may account for this phenomenon. Further, the pH dependence on crosslinking rate (k_2) has yet to be determined. It may initially be assumed that alkaline pH values will have an additive effect on k_2 by increasing the nucleophilicity of the amino acid that crosslinks to the active site,[11] but it can also be suggested that low pH may result in protonation of the epoxide moiety thereby increasing the electrophilic nature and enhancing k_2 as well. Future GT inhibition studies with varying pH will be useful in determining the pH-dependent effects on K_I and k_2. Lastly, it is known[11] that HQS-Epoxy is somewhat competitive in respect to UDP-Glc and RhoA, the concentrations of which are higher than that of biological environments during the *in vitro* assay, and may also confer inhibition of HQS-Epoxy binding to some extent.

Future analysis of HQS-Epoxy *in vivo,* potentially within a mouse model, is needed to determine extent of inhibition, toxicity levels, stability of the peptide, and persistence of inhibitory impact when in a biological environment. Moreover, since HQS-Epoxy portrays a binding affinity and IC_{50} within the high μM range, a future reselection process to identify a mimicking peptide with nM affinity may be beneficial. Determining more accurate values for K_I and k_2 between 0 and 15 minutes, as well as the pH-dependence imposed on these variables, will also further characterize the epoxide-containing peptide and in doing so, these studies will help dictate the potential of HQS-Epoxy as a therapeutic agent to combat CDAD.

CHAPTER TWO METHODS

B. meg protoplast generation. All procedures containing *B. meg* were carried out in a Biosafety Level 2 (BL2) lab following standard operating procedures. *B. meg* cells were grown overnight in 5 mL Difco™ Antibiotic Medium 3 (AB3) at 37 °C and 250 rpm. Fresh, warmed AB3 medium (50 mL) was then inoculated with 1 mL of overnight culture and incubated at 37 °C and 250 rpm until an A_{550} of 1.0 was reached. Cells were harvested (2000 rpm, 10 min, ambient temp), resuspended in 5 mL SMMP (1:1 AB3/SMM (40 mM maleic acid, 80 mM NaOH, 40 mM $MgCl_2$, 1 M sucrose)), and transferred to a 100 mL flask before adding 2 mg/mL lysozyme and incubating (37 °C, 100 rpm, 45-60 min). Again, cells were harvested (2250 rpm, 10 min, ambient temp), washed with 5 mL SMMP, and harvested once more as before. Finally, cells were resuspended in 5 mL SMMP (containing 10% (w/v) glycerol) and aliquoted before storing at -80 °C. Protoplasts were visualized by phase-contrast microscopy.

B. meg plasmid transformation. *B. meg* protoplasts (200 µL) were thawed and 0.5-1 µg of *E. coli* miniprepped-plasmid DNA was added, followed by addition of 600 µL PEG-P (40% (w/v) PEG6000, 500 mM sucrose, 20 mM sodium maleinate, 20 mM $MgCl_2$, pH 6.5) solution was inverted to mix. Cells were incubated on ice and then at room temperature for 2 min each, followed by addition of 2 mL SMMP, gentle inversion, and harvesting (3000 rpm, 10 min, ambient temp). Supernatant was removed and 500 µL fresh SMMP was added before outgrowing at 37 °C and 100 rpm for 90 min. Transformed culture was finally mixed with topagar, plated (LB agar plus 10 µg/mL tetracycline (tet)), and incubated at 37 °C for 16 hours. Colonies were restreaked on tet plates to ensure selection.

rTcdA/B protein expression and purification. *B. meg* cells containing the rTcdA/B plasmid were first grown overnight (37 °C, 250 rpm, 16-19 hours) in 10 mL LB containing 10

μg/mL tet before inoculating 1 L LB flasks (10 μg/mL tet) with the 10 mL overnight culture. Cells were then grown to OD_{600} of 0.4, induced by addition of 0.5% D-xylose, and incubated to OD_{600} of 1.5. Cells were harvested (8500g, 10 min, 4 °C) and stored at -80 °C until use. Thawed cell pellets (1-2 L) containing lysis buffer (50 mM Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, pH 8) and protease inhibitors were lysed by sonication (5 cycles at 37% power) and centrifuged to clarify (15,000 rpm, 45 min, 4 °C). Supernatant was filtered through 0.8 μm and 0.22 μm syringe filters, respectively, before removing from BL2 constraint. Crude lysate was purified first via nickel affinity column chromatography, followed by size exclusion column chromatography and a second affinity column to concentrate the protein. Affinity columns were washed with increasing concentrations of imidazole and elution was achieved with 250 mM imidazole. Lysates were filtered through 0.22 μm filters after each chromatography step and preceding dialysis (10,000 MWCO) into storage buffer (50 mM HEPES-K, 100 mM KCl, 1 mM $MgCl_2$, pH 7.5).

rTcdA^{540} protein expression and purification. BL21 DE3 *E. coli* cells containing the rTcdA^{540} plasmid (pNLA 20405;[50] transformed via electroporation) were first grown overnight (37 °C, 250 rpm, 16-19 hours) in 10 mL LB containing 10 μg/mL amp and 34 μg/mL cam before inoculating 1 L LB flasks (10 μg/mL amp and 34 μg/mL cam) with the 10 mL overnight culture. Cells were then grown to OD_{600} of 0.4-0.6, induced by addition of 1 mM IPTG, and expressed 4-5 hours. Cells were harvested (5500g, 10 min, 4 °C) and stored at -80 °C until use. Thawed cell pellets (1-3 L) containing lysis buffer (50 mM Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, pH 8) and protease inhibitors were lysed by sonication (5 cycles at 37% power) and centrifuged to clarify (10,000 rpm, 45-60 min, 4 °C). Supernatant was filtered through 0.8 μm and 0.45 μm syringe filters, respectively, before removing from BL2 constraint. Crude lysate was purified

first via nickel affinity column chromatography, followed by size exclusion column chromatography and a second affinity column to concentrate the protein. Affinity columns were washed with increasing concentrations of imidazole and elution was achieved with 250 mM imidazole. Lysates were filtered through 0.22 μm syringe filters after each chromatography step and preceding dialysis (10,000 MWCO) into storage buffer (50 mM HEPES-K, 100 mM KCl, 1 mM $MgCl_2$, pH 7.5).

RhoA GTPase expression and purification. Rosetta 2 (DE3) *E. coli* cells containing the RhoA-v4 plasmid (pNLA 20504;[50] transformed via electroporation) were first grown overnight (37 °C, 250 rpm, 16-19 hours) in 10 mL LB containing 30 μg/mL kan and 34 μg/mL cam before inoculating 1 L LB flasks (30 μg/mL kan and 34 μg/mL cam) with the 10 mL overnight culture. Cells were then grown to OD_{600} of 0.4-0.6, induced by addition of 1 mM IPTG, and expressed 4-5 hours. Cells were harvested (5500g, 10 min, 4 °C) and stored at -80 °C until use. Thawed cell pellets (1-3 L) containing lysis buffer (50 mM HEPES, 300 mM NaCl, 1 mM $MgCl_2$, pH 8) and protease inhibitors were lysed by sonication (5 cycles at 37% power) and centrifuged to clarify (10,000 rpm, 45-60 min, 4 °C). Supernatant was filtered through 0.8 μm and 0.2 μm syringe filters, respectively. Crude lysate was purified via nickel affinity column chromatography, washed with increasing concentrations of imidazole, and eluted with 150 mM imidazole. Lysates were filtered through a 0.22 μm syringe filter preceding dialysis (3500 MWCO) into storage buffer (50 mM HEPES-K, 100 mM KCl, 1 mM $MgCl_2$, pH 7.5). Protein was concentrated (3500 MWCO) if necessary.

Optical coupled-enzyme assay. Reactions contained 50 μM RhoA, 0.5 mM UDP-Glucose, 35 units/mL pyruvate kinase, 75 units/mL lactate dehydrogenase, 1 mM PEP, and 0.2 mM NADH in storage buffer (50 mM HEPES-K, 100 mM KCl, 1 mM $MgCl_2$, pH 7.5). Toxin

(rTcdA^{540} or rTcdA-WT) at concentrations ranging from 32.5-325 nM, with or without inhibitor present, was added to initiate the reaction (100 µL final volume) at 0, 15, 30 and 60 min time points, and the decline of A_{340} was monitored using an Agilent 8453 UV-VIS spectrophotometer equipped with a circulating water bath set to 37 °C. For inhibition studies, 50, 100, 300, or 600 uM HQS-Epoxy inhibitor peptide was used. Reactions were incubated at 37 °C over the duration of the experiment.

HQSPG$_{allyl}$HHGGGC-NH$_2$ peptide protection, epoxidation and purification. [11] HQSPG$_{allyl}$HHGGGC-NH2 (HQS-Parent) was purchased from American Peptide, resuspended to a final concentration of 1 mg/mL in 20% ACN/H_20, and purified by reverse-phase high-performance liquid chromatography (RP-HPLC) using a semi-preparative C-18 column (5-65% linear gradient of 0.1% TFA/ACN, flow 1.00 ml/min). Peptide fractions were collected, flash-froze via liquid nitrogen, and lyophilized overnight. A 1:1 molar ratio of dried HQS-Parent peptide and DTT was dissolved in 0.1 M $(NH_4)_2CO_3$ buffer pH 8, flushed with N_2 and incubated at 54 °C for 30 min. Next, solid iodoacetamide was added in 10-fold molar excess. Reaction was kept under N_2 in the dark, with continuous stirring for 2 hours, and then purified by RP-HPLC and lyophilized as above. Keeping under N_2, dried HQS-Protected peptide and 10-fold molar excess *m*-chloroperoxybenzoic acid (mCPBA) was added to a stirred solution of 2:1 CH_2Cl_2:0.01 M Na_2HPO_4/NaH_2PO_4 buffer pH 8, and incubated at room temperature with continuous stirring for at least 5 hours. Reaction progress was monitored by thin-layer chromatography (mobile phase 7:2:1 NH_4OH: H_20: 2-propanol). Peptide was precipitated and washed 6 times by addition of diethyl ether to the aqueous layer. Peptide was then lyophilized, redissolved in water, and desalted by addition of ACN (50% v/v), centrifugation (4 °C, 10,000g, 10 min), and removal of supernatant from salt pellet. Again, peptide was purified as above, lyophilized, and stored at -20

°C until use. Mass spectrometry after each step determined presence of peptide.

Vero cell cultures. All procedures containing Vero cells were carried out in a Biosafety Level 2 (BL2) lab following standard operating procedures. Vero cell glycerol stocks were removed from -80 °C and quickly (less than 1 min) warmed to 37 °C in a water bath. Cells were suspended in 6 mL warm Eagle's Minimum Essential Media (EMEM) enriched with 10% fetal bovine serum (FBS) and 1% Antibiotic-Antimycotic (Anti-Anti) and spun down (200g, 5 min, ambient temp). Media was removed from pellet and cells were resuspended in 3 mL fresh media, which was plated in a 25 mL culture flask and incubated at 37 °C and 5 % CO_2. Once grown to roughly 80% confluency, adherent cells were passaged and regenerated as followed: EMEM was removed and 0.5 mg/mL trypsin was added before incubating the flasks at 37 °C and 5% CO_2 for 3 minutes or less. Trypsinized cells were then diluted 3X with EMEM, harvested (200g, 5 min, ambient temp), resuspended with 5 mL EMEM and new culture flasks were seeded with 50 µL resuspended cells and 3 mL fresh EMEM before incubating as before. Newly plated cells from glycerol stocks underwent a minimum of three passages before being used in any experiments.

Visual cell protection assay. Vero cells were seeded in a clear, 12-well plate in 500 µL Eagle's Minimum Essential Media (EMEM) enriched with 10% fetal bovine serum (FBS) and 1% Antibiotic-Antimycotic (Anti-Anti), incubated at 37 °C and 5 % CO_2, and grown to roughly 80% confluency. Wells were imaged using a Nikon Eclipse TS100 (Nikon Intensilight C-HGFI) and washed twice with 1 mL serum-free EMEM (SF-EMEM supplemented with 1% Anti-Anti) before adding either SF-EMEM alone, 0.2 nM rTcdA-WT, 0.2 nM rTcdA-WT plus 600 µM HQS-Epoxy peptide inhibitor, or 600 µM HQS-Epoxy peptide inhibitor topped off with SF-EMEM to a final volume of 500 µL. Cells were then incubated (37 °C, 5 % CO_2) for 1 hour before imaging again.

CHAPTER THREE:

Glucosylation Drives the Innate Inflammatory Response to *Clostridium difficile* Toxin A[38]

In effort to further understand the role of immune system activation in CDAD, the Feig Lab collaborated with the Petri Lab in the Immunology Department at the University of Virginia. The Petri Lab has recently reported evidence of elevated levels of specific proinflammatory cytokines (IL-23 and IL-1β) that enhance disease progression in specimens positive for *C. diff* infection.[37, 53] Furthermore, Cowardin, *et al.,* allude to the fact that activation of a robust inflammatory response is the result of not only *C. diff* toxins, but also the presence of a necessary precursor, or "priming" signal, brought on by bacterial components known as pathogen-associated molecular patterns (PAMPs).[37] Since part one of my thesis project deals solely with inhibition of GTD, the recent discoveries from the Petri Lab are of particular interest to me, not to mention the fact that they have access to mouse models. Moreover, the experiments in the Petri Lab, until now, were done with naturally isolated *C. diff* toxins, but that precludes the use of modified and mutant toxins, which provide a more granular probe of toxin activity. With the Feig Lab's background in biochemistry, the Petri Lab's knowledge of immunology, and the common interest of CDAD, a collaborative partnership was formed. Goals of this collaboration included further mechanistic understanding of inflammasome activation upon *C. diff* exposure, specific pathways involved in inflammation upregulation, and specific players associated with these pathways, which can be seen in **Figure 11**.

A previous lab member, Dr. Amy Kerzmann, genetically modified wild type rTcdA by site-directed mutagenesis to produce an enzymatically-inactive mutant, named rTcdA-NVN, with commonly known D285N and D287N mutations.[50, 54] To begin collaboration with Carrie Cowardin in the Petri Lab, expression plasmids containing these genes (rTcdA-WT,

pANK_80406; rTcdA-NVN, pANK_80407) were transformed into *Bacillus megaterium* (*B. meg)*, expressed, and purified before shipping to the University of Virginia. Final SDS-PAGE gels depicting the purity of these proteins are shown in **Figure 12A**. The final elution fraction (E*) containing the toxins also shows the presence of a protein band just below rTcdA that was unable to be resolved and separated with the size-exclusion column in our lab due to very similar molecular weights. Also existing in the E* lane is a band below 72 kDa, which was previously identified by mass spectrometry[11] to be the 63 kDa GTD that is produced by self-cleavage of rTcdA at the CPD. The higher molecular weight positioning on the gel is most likely due to different charge distribution within the molecules. Although not yet tested by mass spectrometry, it would seem apparent then that the small band above 170 kDa and below rTcdA is likely rTcdA with the loss of GTD (roughly 245 kDa) after self-cleavage.

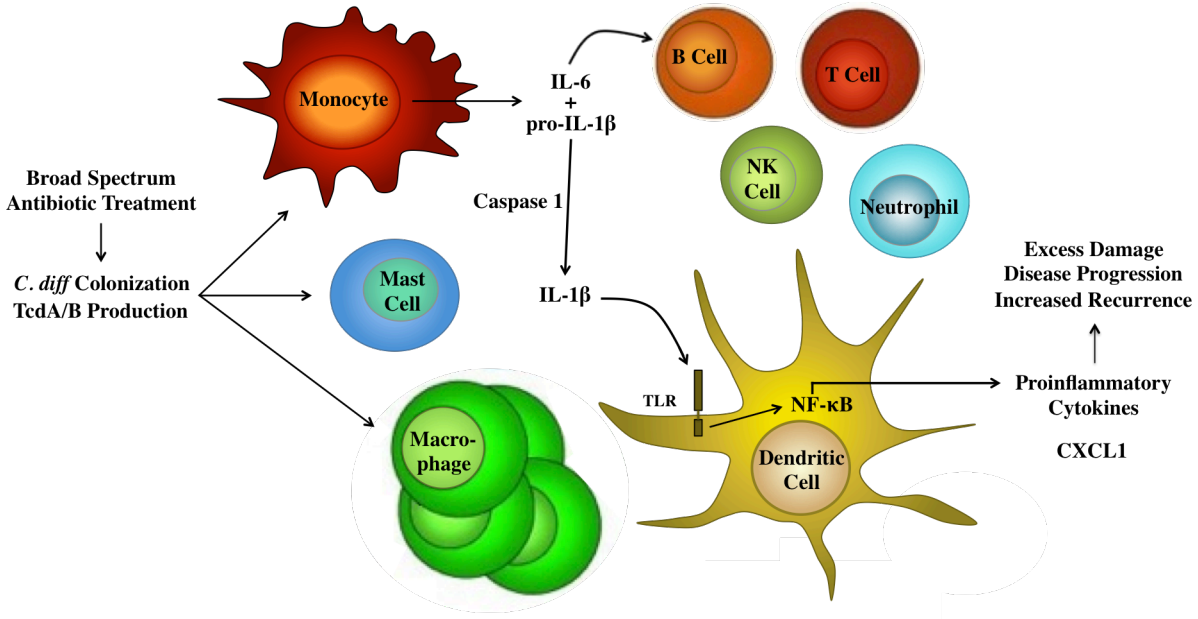


Figure 11: Overview and key components of *C. diff* immune system activation pathways. *C. diff* PAMPs and toxins produced after colonization activate the host innate immune system by initiating cytokine production and recruiting immune cells (macrophages, mast cells, monocytes, etc.) to the site of infection. Immune cells subsequently express more cytokines, creating a positive feedback loop. The proinflammatory cytokines IL-6 and pro-IL-1β are produced mainly by monocytes; IL-6 results in the recruitment of other immune cells (B cells, T cells, natural killer (NK) cells, and neutrophils) while pro-IL-1β becomes IL-1β after caspase 1 cleavage.[55, 56] IL-1β then binds to dendritic cell surface toll-like receptors (TLR) where it initiates expression of the NF-κB protein complex, a transcriptional regulator, which in turn upregulates proinflammatory cytokines, such as CXCL1.[57, 58, 59]

rTcdA-WT and NVN mutant were assayed for activity by the CellTiter-Glo® Luminescent Cell Viability Assay.[60] Data in **Figure 12B** shows decreasing ATP concentrations (shown as relative light units; RLU) in the presence of increasing rTcdA-WT, indicating functional glucosyltransferase activity and loss of cell viability, where the cell population was killed with a half-maximal inhibitory concentration (IC_{50}) near 0.7 nM. Cell viability is also reduced with increasing rTcdA-NVN, however cell death is significantly less than that seen for rTcdA-WT treatment, and is most likely due to toxicity from leaky endosomes since pore formation is still occurring. Furthermore, increasing the concentration of toxin results in less volume of media per well, which may also contribute to cell death. Although utilization of a wider concentration range of these toxins would have yielded more valuable data in **Figure 12B**, a visual verification of rTcdA-WT activity can also be seen in **Figure 13**. Phase-contrast microscopy displays rounding of HeLa cells after a 24-hour incubation period with 0.2 nM rTcdA-WT, indicating apoptosis and cell death.

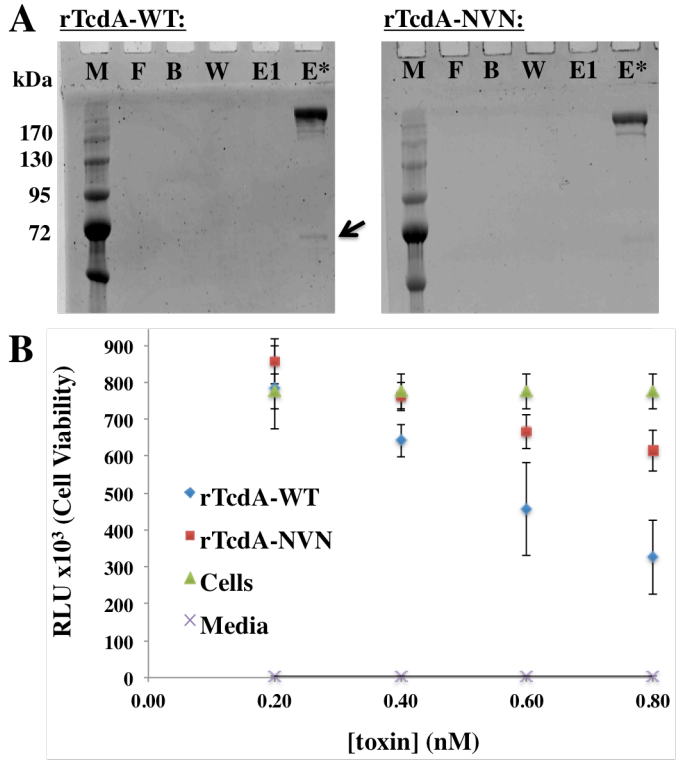


Figure 12: Purity and activity of rTcdA-WT and rTcdA-NVN. A: SDS-PAGE gels after purification of rTcdA-WT (left) and NVN (right). M: protein marker, F: column flowthrough, B: wash fraction (20 mM imidazole), W: wash fraction (50 mM imidazole), E1: first elution fraction (0.5 mL, 250 mM imidazole), E*: pooled elution fractions containing rTcdA (2.5 mL). Protein marker molecular weights are shown to the left and the arrows indicates the self-cleavage products. **B:** Graph of cell viability data after incubation with toxins. RLU: relative light units. Blue diamonds: rTcdA-WT; red squares: rTcdA-NVN; green triangles: cells with no toxin; purple crosses: media background with no cells. n = 9.

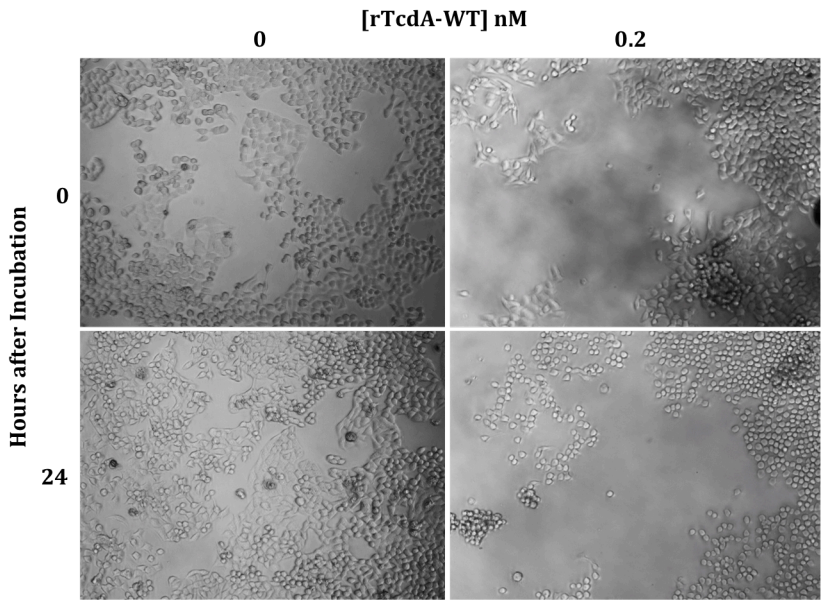


Figure 13: Phase-contrast microscopy images of rTcdA-WT-treated HeLa cells. HeLa cells incubated with 0.2 nM rTcdA-WT (right) were imaged before (top) and after 24 hours (bottom). Control cells (left) were not treated with toxin.

Using these recombinant toxins in Dr. Bill Petri's lab, *in vitro* experiments done by Dr. Carrie Cowardin[38] with murine bone marrow derived dendritic cells (mBMDCs) and human THP-1 monocytes (hTHP-1)[37] indicated significantly elevated levels of CXCL1, IL-6, IL-1β, and caspase 1 after incubation with rTcdA-WT in comparison to NVN and controls (**Figure 14A-B**). CXCL1, also known as growth-related oncogene alpha or chemokine (C-X-C motif) ligand 1, is a chemokine (i.e., chemotactic cytokine) that uses methods of chemotaxis to recruit leukocytes to the site of inflammation and infection.[57] IL-1β and IL-6, cytokines predominantly produced by monocytes, promote damaging inflammation via cellular infiltration. IL-1β is responsible for the induction of IL-6, as well as many other cytokines, and is even involved in a positive feedback loop for its own upregulation. IL-1β's major contribution to inflammation comes from activation of the arachidonate/eicosanoid metabolism. Although IL-1β and IL-6 share many biological roles, IL-6 is further involved in stimulating B-lymphocyte maturation and T-cell regulation, particularly the differentiation of T-cells into cytotoxic T-cells.[55] Caspase 1, also named interleukin-1 converting enzyme (ICE), is responsible for the proteolytic cleavage of pro-IL-1β to its active IL-1β form.[56] **Figure 14B** exhibits increased levels

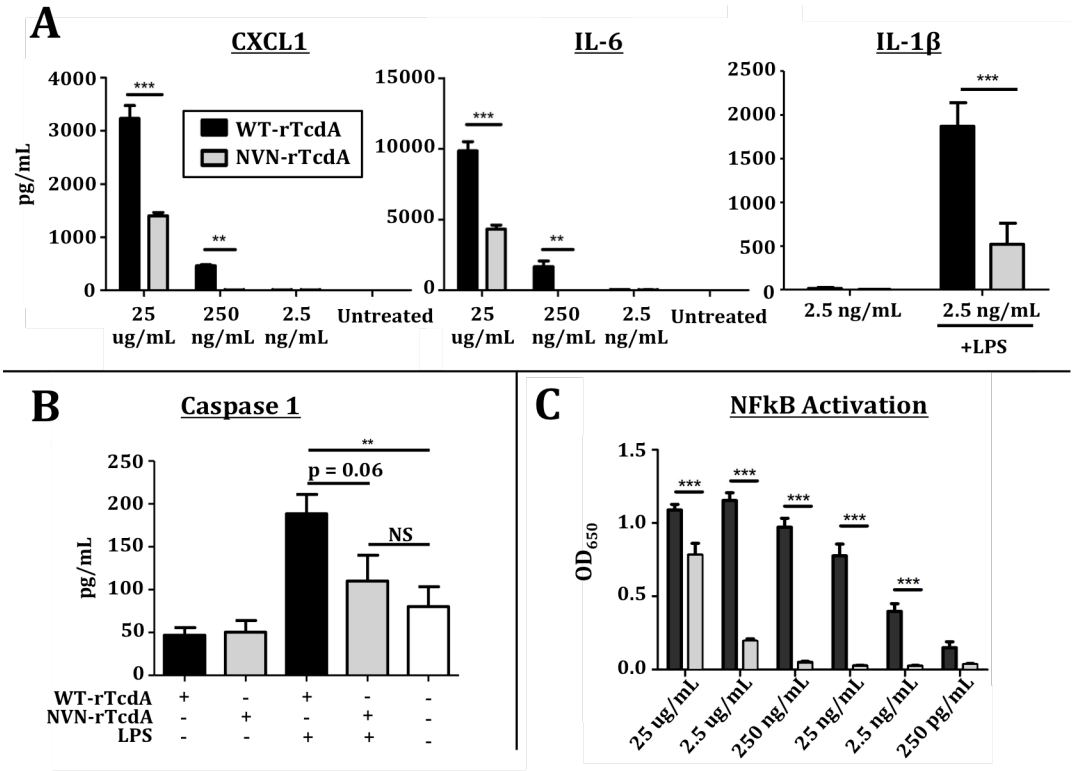


Figure 14: rTcdA-WT activates inflammasome and cytokine production *in vitro*. mBMDCs **(A)** and hTHP-1 monocytes **(B)** were incubated with indicated concentrations of rTcdA toxins plus or minus LPS priming. Lysates were analyzed by ELISA for protein production after 24 hours. **C:** Murine RAW-Blue macrophages were subject to displayed rTcdA concentrations. SEAP levels reporting NF-κB activation were quantified in the culture media by the QUANTI-Blue assay at 24 hours.[38]

of caspase 1, indicating a systemic inflammatory response upstream of IL-1β. Moreover, the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-ϰB) pathway, which is a blueprint of proinflammatory signaling, shows to be significantly triggered in the presence of rTcdA-WT (**Figure 14C**). Activated by cytokines, such as IL-1 and TNFα, this pathway involves several proteins acting as regulatory transcriptional factors for genes encoding proinflammatory cytokines, chemokines, and adhesion molecules, including CXCL1 mentioned above.[58, 59] These data (recently published)[38] indicate that glucosyltransferase activity is needed to induce such inflammatory responses, and in the case of caspase 1 and IL-1β, a priming signal is needed as well.

Also done in the Petri Lab by Dr. Carrie Cowardin,[38] injections of rTcdA toxin were

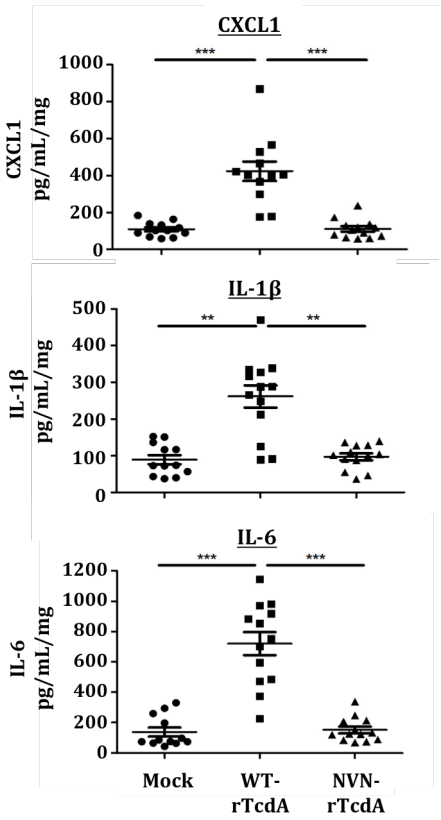


Figure 15: rTcdA-WT initiates cytokine production *in vivo*. Indicated toxin (15 μg) was injected into the cecum of mice. After 8 hours, mice were sacrificed and cecal tissue lysates were analyzed for cytokine levels by ELISA. Each data point represents an individual mouse.[38]

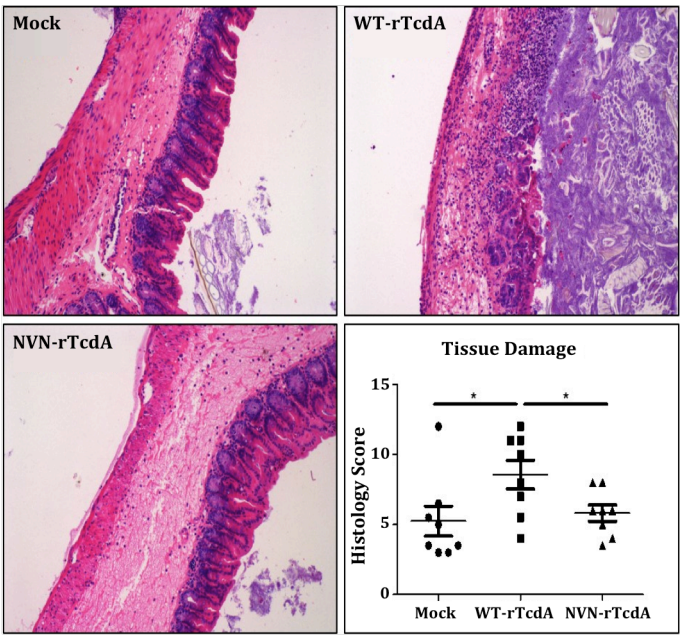


Figure 16: Histology report of mice cecal sections in response to *C. diff* rTcdA. Murine cecal sections were analyzed by hematoxylin-eosin staining after treatment with WT-rTcdA (top right) or NVN-rTcdA (bottom left). Control mice (Mock; top left) were not subject to either toxin. Quantitation of tissue damage for each group of mice is shown in the bottom right, where each data point represents one individual mouse (n = 8/group).[38]

placed into the cecum of mice by laparotomy and subsequently subjected to cecal tissue analysis. **Figure 15** demonstrates the escalated appearance of cytokines (CXCL1, IL-1β, and IL-6) determined by ELISA, within the *in vivo* surroundings of the mouse colon. The destruction caused by rTcdA-WT and said proinflammatory cytokines can be visualized in **Figure 16**. Mice injected with rTcdA-WT show a significantly disrupted epithelial barrier in comparison to rTcdA-NVN and controls.

The growing issue and prevalence of CDAD has shown to be more complicated than previously known. Not only does the *C. diff* pathogen play a major role in the disease, but recent evidence also shows the host immune system is partially to blame.[32, 37, 53] These data demonstrate that the host immune system binds to and recognizes *C. diff* PAMPs as part of the priming

process. However glucosylation of cellular targets by rTcdA is necessary to induce significant inflammation and cytokine production.[38] These results further our understanding of how the *C. diff* glucosylating toxins and the host immune system act in concert to mediate CDAD colitis. The outcomes lend support to the ongoing development of toxin inhibitors, as this work shows that if delivered and dosed correctly, inhibited toxins would not activate the immune system in a manner that would damage the colon. This would also provide time for the patient to complete their course of primary antibiotics and undergo fecal transplantation to restore the gut microbiome to its initial state prior to the antibiotic therapy that allowed *C. diff* to colonize the patient's GI tract.

CHAPTER THREE METHODS

HeLa cell cultures. All procedures containing HeLa cells were carried out in a Biosafety Level 2 (BL2) lab following standard operating procedures. HeLa cells were initially obtained from the Pflum Lab at Wayne State University. Cells were grown in 25 mL culture flasks containing Eagle's Minimum Essential Media (EMEM) enriched with 10% fetal bovine serum (FBS) and 1% Antibiotic-Antimycotic (Anti-Anti) and incubated at 37 °C and 5 % CO_2. Once grown to roughly 80% confluency, adherent cells were passaged and regenerated as followed: EMEM was removed and 0.5 mg/mL trypsin was added before incubating the flasks at 37 °C and 5% CO_2 for 3 minutes. Trypsinized cells were then diluted 3X with EMEM and harvested (1000g, 7 min, ambient temp). Cells were resuspended with 5 mL EMEM and new culture flasks were seeded with 50 μL resuspended cells and 3 mL fresh EMEM before incubating as before.

Luminescent cell viability assay. CellTiter-Glo® luminescent reagents were purchased from Promega and the protocol was performed as previously described.[60] Briefly, HeLa cells

(30,000 cells/well) were cultured in an opaque 96-well plate with fresh EMEM containing 10% FBS and 1% Anti-Anti and incubated overnight at 37 °C and 5% CO_2. Cells were washed with serum-free EMEM (SF-EMEM) before addition of fresh SF-EMEM and varying concentrations of toxins to a total volume of 50 μL per well. Next, cells were incubated as above for 48 hours before cooling to room temperature, washing as above, and replacing fresh SF-EMEM (50-100 μL) in wells. An equal volume of CellTiter-Glo® luminescent reagent was added, the plate was agitated (2 min, ambient temp), and then luminescent data was acquired using a GENios Tecan Microplate Reader.

In vitro inflammasome activation and cytokine expression level determination. Carrie Cowardin, a collaborator in the Petri Lab at the University of Virginia, performed these experiments with rTcdA toxin provided by the Feig Lab (data recently published).[38] Briefly, cultured mammalian cells (murine bone marrow derived dendritic cells, human THP-1 monocytes, murine raw macrophages) were incubated for 24 hours with various concentrations of rTcdA-WT, rTcdA-NVN mutant, or neither (control), and in some cases lipopolysaccharide (LPS) as a priming signal. Cells were lysed and subject to ELISA to establish protein concentrations (CXCL1, IL-6, IL-1β, and Caspase 1). Quantitation of NF-κB upregulation i.e., inflammasome activation, was carried out with murine Raw-Blue™ macrophages, which are modified to contain a secreted embryonic alkaline phosphatase (SEAP)-reporter construct that is inducible by NF-κB. Culture media was surveyed by the QUANTI-Blue™ assay at A_{650}[61, 62] after 24 hours.

In vivo cytokine expression level determination. Carrie Cowardin from the University of Virginia performed these experiments.[38] In brief, mice underwent a cecal laparotomy;[63] meaning rTcdA-WT and rTcdA-NVN toxins were injected into the cecum (15 μg). After 8 hours, mice

were sacrificed. Cecal tissues were obtained, lysed, and the supernatant was subject to ELISA in order to quantitate cytokine production.

Immunohistochemistry staining and histological analysis. Carrie Cowardin (University of Virginia) also performed these experiments. Again, toxin was injected to the cecum of mice via laparotomy. Mice were then sacrificed after 8 hours as above. Cecal tissues were obtained, fixed with Bouin's solution (picric acid, acetic acid, formaldehyde) for 18 hours, embedded with paraffin for rigidity, sectioned, and stained with hematoxylin and eosin to visualize nuclei and the rest of the cell components, respectively.[37, 38] Histology score was based on 5 parameters: Submucosal edema, inflammatory infiltrate, epithelial disruption, luminal exudate, and mucosal thickening.

CHAPTER FOUR:

Cloning, Expression, and Purification of *Clostridium difficile* Recombinant Toxin B

The homologous *C. diff* enterotoxins A and B share 48% sequence identity,[64] and as seen in **Figure 3**, the main structural difference between the two resides within the length of the CROP domain. Although TcdA/B exhibit similar mechanisms of infection and enzymatic functions, TcdB is actually 100-fold more efficacious at initiating cytopathic cell rounding and apoptotic cell death,[41] most likely attributed to a shorter CROP domain that is less able to occlude GTD and CPD activity.[16, 48] Also in contrast to TcdA, high concentrations of TcdB have been shown to cause necrotic cell death via an alternative NADPH oxidase-mediated pathway, portraying the sudden depletion of cellular ATP, release of lactate dehydrogenase, and disruption of the cell membrane.[41, 65, 66] Interestingly, the necrosis phenotype has been revealed to persist in cells treated with CPD/GTD-inactive TcdB mutants,[65] indicating the sole necessity for functioning TD[39, 64] and CROP domains.[41] To further advocate the potential importance of TcdB-induced necrosis regarding CDAD, studies show the elevated TcdB concentrations needed to exhibit this phenotype not only correlate to pathological colonic epithelial damage, but also compare to levels found in feces of clinical isolates with severe *C. diff* infection.[65, 67] Lastly, there are numerous hypervirulent and resistant *C. diff* strains emerging today, such as the early 2000s epidemic strain NAPI/027/BI that is known for increased TcdA/B production (roughly 16-fold and 23-fold, respectively) and ribotype 017, which is a sole producer of TcdB (A^-B^+).[1, 68, 69] Because of the numerous aforementioned reasons, a more complete analysis of *C. diff* TcdB and therapeutic methods that successfully target this toxin are clinically important, and as all previous studies done in the Feig Lab have utilized *C. diff* TcdA, the cloning, expression, and purification of toxin B was needed.

Considering the great length of the TcdA/B genes, previous methods to clone rTcdA-WT and rTcdA-NVN in the Feig Lab included splitting the gene into segments to avoid possible errors in replication.[53] Although this technique was successful, it is also strenuous and time consuming. To sidestep previous methods, the DNA polymerase used to clone rTcdB was upgraded to Q5® High-Fidelity polymerase, which has 3' to 5' exonuclease activity and an error rate of roughly 100-fold lower than *taq* and 12-fold lower than *pfu*.[70] Q5® polymerizes at an extremely fast rate of 6 kb/min and has shown to successfully replicate amplicons of less than or equal to 20 kb,[71] which is more than 2.8 times the length of the targeted 7101 bp *tcdb* gene. Furthermore, Q5® polymerase results in blunt ends, allowing for uncomplicated TOPO cloning, is effective at manipulating DNA via site-directed mutagenesis,[71] and has proven here to clone the entire *tcdb* gene without error.

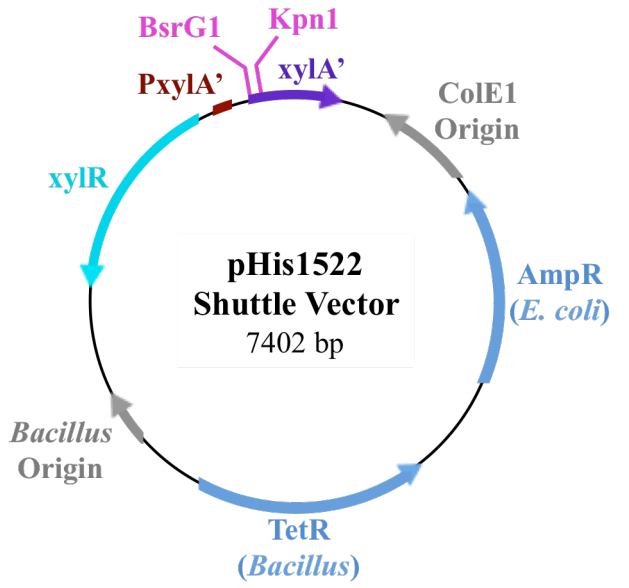


Figure 17: pHis1522 shuttle vector for *B. meg* expression system. The pHis1522 plasmid allows for ease of use between *E. coli* and *B. meg*, containing specific replication origins (grey) and antibiotic resistance genes (light blue) for each. Downstream of the xylose isomerase promoter (PxylA', dark red), the multiple cloning site (magenta) is positioned in frame with the xylose isomerase gene (xylA', purple) with specific BsrG1 and Kpn1 restriction cut sites shown in pink. Finally, the transcriptional regulatory repressor of xylose isomerase (xylR) is shown in cyan.

Full length TcdA/B proteins are rather large (308 kDa and 270 kDa, respectively) and do not express well in *E. coli*, therefore the *Bacillus megaterium* (*B. meg*) expression system (MoBiTec)[72] known for oversized protein expression was used to provide a route for holotoxin production. The pHis1522 plasmid utilized in this system (**Figure 17**), possessing *E. coli*-specific ampicillin and *B. meg*-specific tetracycline resistance markers, is a

shuttle vector with relieved codon bias that allows for DNA manipulation in *E. coli* before transforming into *B. meg* for expression. Furthermore, the multiple cloning site within pHis1522 contains a built-in C-terminal 6X his tag and is positioned within xylA' (xylose isomerase-encoding gene) to permit xylose induction of TcdA/B protein expression.[72]

A general overview of the *C. diff tcdb* cloning process, borrowed closely from Feng, *et al.*,[73] is illustrated in **Figure 18**. To begin, *tcdb* from *C. diff* genomic DNA (strain ATTC 9689D) was isolated and amplified by PCR utilizing Q5® High-Fidelity polymerase and primers 1 and 2

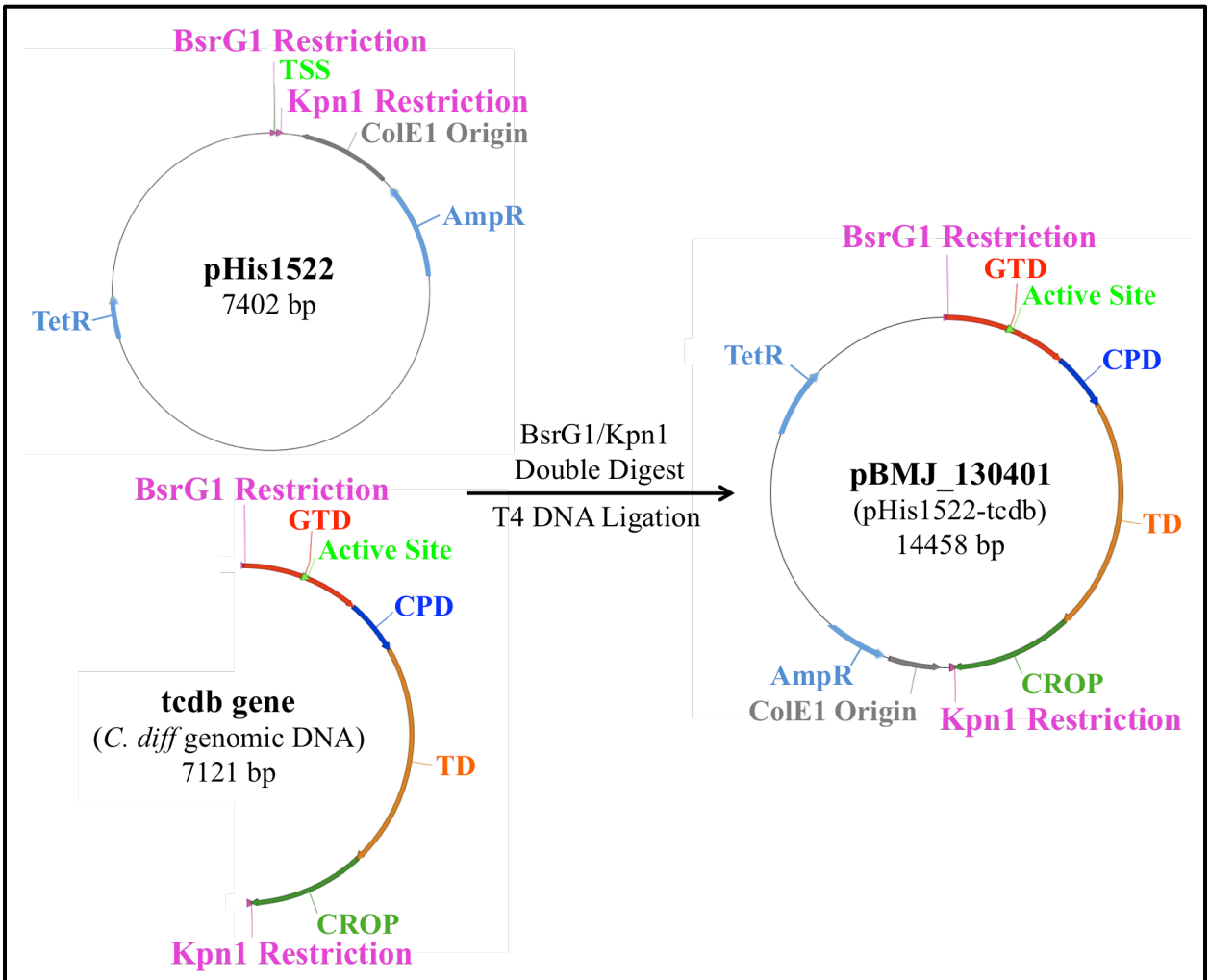


Figure 18: Overview of rTcdB-WT cloning process. The 7121 bp *tcdb* gene was isolated from *C. diff* genomic DNA (strain ATTC 9689D) and amplified via Q5® High-Fidelity DNA polymerase using primers 1 and 2, which were designed to contain 5' BsrG1 and 3' Kpn1 restriction digest sites for convenient ligation into pHis1522 (7402 bp). Both *tcdb* and pHis1522 were purified, digested with BsrG1 and Kpn1 enzymes, and ligated together using T4 DNA ligase to give the final expression plasmid pBMJ_130401 (14458 bp). TSS: transcriptional start site for the xylA' gene.

Table 2: ***Primers for cloning rTcdB-WT and rTcdB-NVN.***

Primer Number	Name	Primer Sequences
1	**TcdB_F**	5' - GCGC**TGTACA**ATGAGTTTAGTTAATAGAAAAC - 3'
2	**TcdB_R**	5' - ATATAT**GGTACC**CTTCACTAATCACTAATTGAGC - 3'
3	**B2R_2**	5' - TTAATCTTAGGTCTATCAGAAATTATACTTG - 3'
4	**rTcdB_NVN_F2**	5' - GAAATTGGTGGTATGTATTTAaATGTTaATATGTTACCAGGAATACAACCAGAC - 3'
5	**rTcdB_NVN_R2**	5' - GTCTGGTTGTATTCCTGGTAACATATtAACATtTAAATACATACCACCAATTTC - 3'

NOTE: Bold: restriction enzyme site. Underlined: primer binding complementary site. Lower case: uncomplementary bases for SDM.

(**Table 2**), which were equipped with 5' BsrG1 (primer 1; forward) and 3' Kpn1 (primer 2; reverse) restriction digest sites for convenient ligation into the multiple cloning site of the pHis1522 shuttle vector. Following PCR, the expected amplicon size of 7121 bp (full *tcdb* gene length plus primer overhang) was confirmed by agarose gel electrophoresis and is shown in **Figure 19**, where all attempted temperatures within the gradient ranging from 49 °C to 61 °C were effective for primer annealing and 53.4 °C was used for subsequent scale-up PCR reactions.

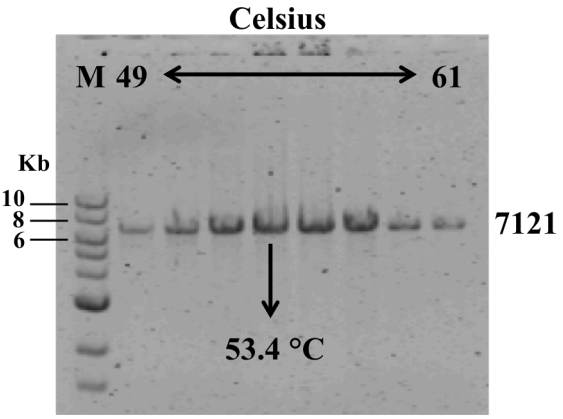


Figure 19: The *tcdb* gene was successfully isolated from *C. diff* genomic DNA. Agarose gel shows successful amplification of the 7121 bp *tcdb* amplicon. All temperatures ranging from 49 °C-61 °C resulted in effective annealing and 53.4 °C was used for following scale-up PCR reactions.

Once clean and quantified by A_{260}, the *C. diff tcdb* amplicon and the shuttle vector pHis1522 were digested with BsrG1 and Kpn1 restriction enzymes, ligated together using T4 DNA Ligase, and transformed into Top10 electrocompetent *E. coli* cells by electroporation. The ampicillin resistance marker within pHis1522 was used to select for transformant colonies containing viable plasmid, five of which were then isolated and purified. To confirm plasmid length prior to sequencing, Xho1 and Spe1 restriction enzymes with plasmid-specific and toxin-specific cut sites, respectively, were used to confirm the correctly sized products of 8647 and

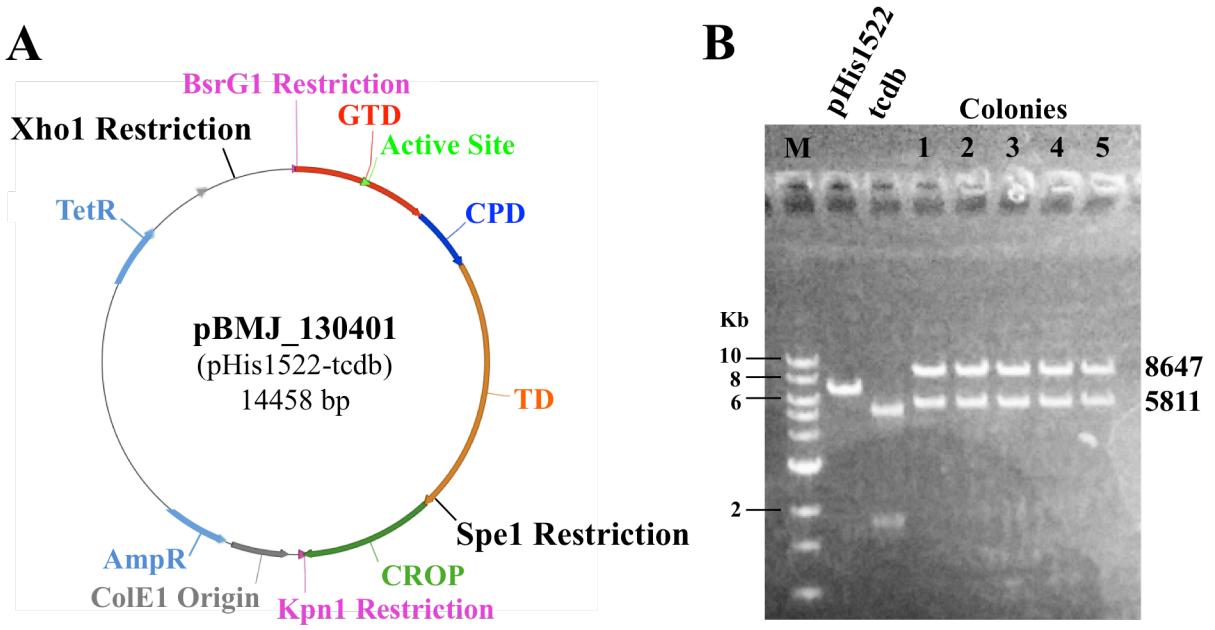


Figure 20: Double digest verification for pBMJ_130401. A: Plasmid map of pBMJ_130401 depicts the *tcdb* gene insert and Xho1/Spe1 digest sites used for determining correct plasmid size. **B:** DNA isolated from five different colonies showed the expected products of 8647 and 5811 bp after Xho1/Spe1 digestion.

5811 bp (**Figure 20**). Primers shown in **Table 3** subsequently verified the 7101 bp *tcdb* gene sequence to contain no errors. Finally, the tcdb-pHis1522 plasmid, now renamed pBMJ_130401, underwent transformation into *B. meg* for protein production. **Figure 21** shows purified rTcdB-WT at the correct size of 270 kDa, where a 16 hour expression period[73] of 1 liter cells resulted in roughly 1 mg of protein, which is roughly twice as much as toxin A per liter and most likely due to protein length. Furthermore, incubation of HeLa cells with 0.2 nM rTcdB-WT caused cytopathic rounding effects within a 2 hour period, proving the rTcdB-WT protein to be active (**Figure 21**).

For negative control, just as was done for rTcdA-NVN,[53] creation of the enzymatically inactive mutant rTcdB-NVN is dependent upon the aspartate to asparagine mutation within the active site of the GTD (D286N, D288N). To avoid producing error in the rest of the *tcdb* gene, toxin B GTD was first isolated and PCR amplified from pBMJ_130401 (1946 bp, **Figure 22A**) using primers 1 and 3 from **Table 2**, followed by direct ligation into TOPO vector

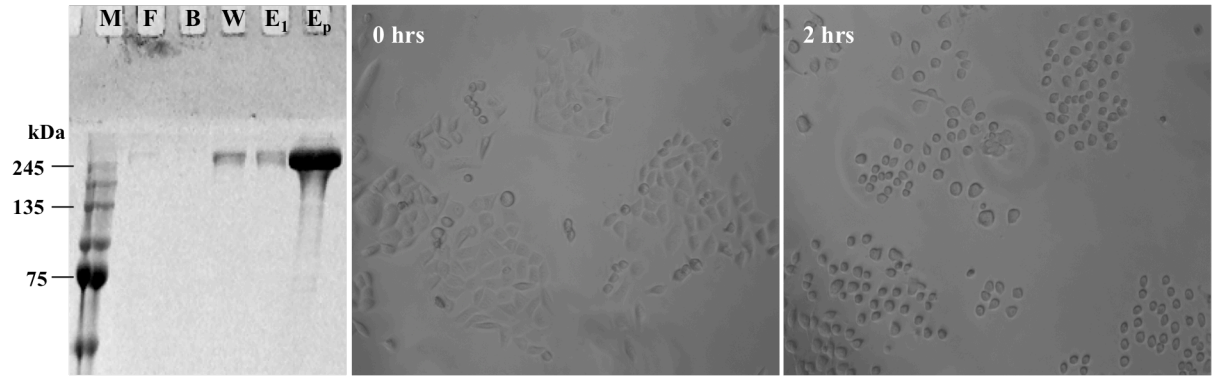
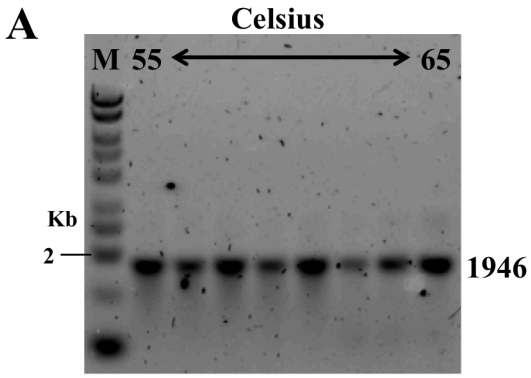
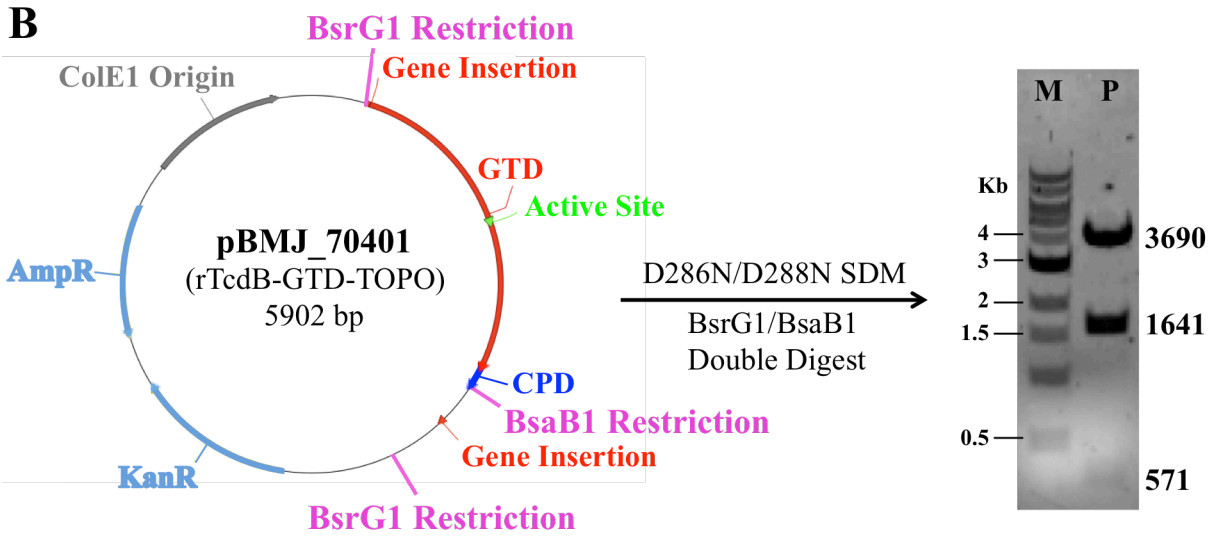


Figure 21: SDS-PAGE gel of purified rTcdB-WT and cytopathic rounding effects of active protein. Left: A clean band in the E_p (elutions pooled) lane shows the expected size of 270 kDa. M: marker. F: flowthrough. B: binding buffer wash 1. W: wash buffer wash 2. E_1: elution 1. Incubation of healthy HeLa cells (0 hours, **middle**) with 0.2 nM rTcdB-WT for 2 hours resulted in characteristic cytopathic rounding effects (**right**).

Figure 22: Isolation and site-directed mutagenesis of rTcdB-GTD. A: rTcdB-GTD was isolated and amplified by PCR from pBMJ_130401 using primers 1 and 3 in **Table 1**. All annealing temperatures between 55°C and 65°C resulted in the expected amplicon length of 1946 bp. **B:** The GTD was then ligated into TOPO (pBMJ_70401; 5902 bp). Following transformation into *E. coli* cells and kanamycin selection, the DNA from resultant colonies underwent SDM using primers 4 and 5 to create D286N and D288N mutations. Plasmid lengths were verified by digestion with the BsrG1 and BsaB1 enzymes, to produce digest products of 3690, 1641, and 571 bp. M: marker. P: digestion reaction products.





(pBMJ_70401; **Figure 22B left**), transformation into electrocompetent *E. coli* cells, and plating on kanamycin antibiotic for selection. Again, select transformant colonies were grown overnight to produce ample amounts of DNA that was then isolated and purified. Site-directed mutagenesis via PCR with primers 4 and 5 (**Table 2**) was done to confer D286N and D288N mutations within the GTD (pBMJ_70402), and digestion of the PCR product with BsrG1 and BsaB1 restriction enzymes resulted in products of 3690, 1641, and 571 bp which confirmed correct plasmid length as seen in **Figure 22B**. Although the smallest digest product of 571 bp ran below the indicated 500 bp marker, the first two bands and the overall digestion pattern within the gel seemed to be correct, therefore the plasmid was sent forth for sequencing and verified via the primers in **Table 4**. In order to create the final rTcdB-NVN *B. meg* expression plasmid, pBMJ_70402 and pBMJ_130401 were digested with BsrG1 and BsaB1 to give the rTcdB-GTD-NVN (1641 bp) and pHis1522-*tcdb* minus the GTD (12817 bp) products, which will need to undergo ligation and transformation into *B. meg*.

For many reasons, comprehending toxin B functionality in comparison to toxin A, as well as determining inhibitory treatment methods that are specified to each toxin is clinically important for impeding CDAD as a whole. Not only do some hypervirulent strains express toxin B as the sole virulence factor, but also TcdB is more potent in comparison to TcdA and it contains a poorly understood pathogenic mechanism that results in necrotic epithelial cell death in patients with CDAD. Experimentation pertaining to TcdA has been underway for quite some time in the Feig Lab, but previous research utilizing TcdB has been very minimal since *C. diff* toxins are rather costly and until now there wasn't an obtainable TcdB expression plasmid. Although the final ligation and transformation of the rTcdB-NVN plasmid have not been completed at this point, the data presented here shows sequence verification for rTcdB-WT

(pBMJ_130401; **Appendix I**) and the components of rTcdB-NVN (**Appendix II**), where expression of rTcdB-WT resulted in ample amounts of active protein that will elicit further mechanistic and inhibitory analysis of *C. diff* toxin B in the future.

Table 3: ***Sequencing primers for rTcdB-WT.***

Primer Number	Name	Sequences of rTcdB-WT Sequencing Primers
1	**B1F**	5' - GGATATTCTGATAGAGAAAT - 3'
2	**B1R**	5' - TCACTATTTACATCTTTC - 3'
3	**B2F**	5' - GAAGAATATCATAATATGTC - 3'
4	**B2R**	5' - TTAATCTTAGGTCTATCAGA - 3'
5	**B3F**	5' - TTACAAGGAGATAAAATTAG - 3'
6	**B3R**	5' - TACTAACTTACCATTTACAG - 3'
7	**B4F**	5' - GAATTTAAACTAATAGAATC - 3'
8	**B4R**	5' - TCTTCTTTTTGTACTTCCAA - 3'
9	**B5F**	5' - TACTTTTAGTTCCTTTAGCA - 3'
10	**B5R**	5' - TCAGATACAAATAAACCTTC - 3'
11	**B6F**	5' - GTTGATTTATTATCTAAATC - 3'
12	**B6R**	5' - GTACTAAATACACCTGTTTG - 3'
13	**B7F**	5' - TAATAACTTTGGAATGATGG - 3'
14	**B7R**	5' - TCACTTTCATTTTCCATATC - 3'
15	**B8F**	5' - GGAAATGAAGAAGGTGAAG - 3'
16	**B8R**	5' - ACTTTGTAAGAATGTCAACT - 3'

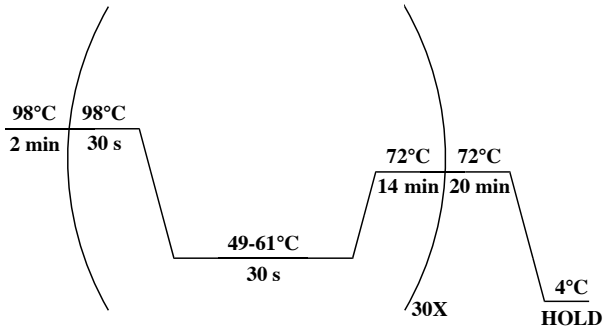
Table 4: ***Sequencing primers for rTcdB-NVN.***

Primer Number	Name	Sequences of rTcdB-NVN Sequencing Primers
1	**B1R**	5' - TCACTATTTACATCTTTC - 3'
2	**B2F**	5' - GAAGAATATCATAATATGTC - 3'
3	**B2R**	5' - TTAATCTTAGGTCTATCAGA - 3'
4	**B3F**	5' - TTACAAGGAGATAAAATTAG - 3'
5	**M13 F (-20)**	standard for TOPO - supplied by sequencing agency
6	**M13 R**	standard for TOPO - supplied by sequencing agency

CHAPTER FOUR METHODS

Clostridium difficile rTcdB-WT cloning. Procedure to isolate, clone, and express *C. diff* toxin B was closely borrowed from Feng, *et al*.[73] Here, *C. diff* genomic DNA (90556-M6S) was purchased from ATTC™ (item number 9689D-5 DR), all primers were synthesized and purchased from Sigma, and PCR reactions were done using a Bio-Rad MyCycler Thermo Cycler (catalog #19209590). In short, *rtcdb* was isolated using primers 1 and 2 (**Table 2**) from *C. diff* genomic DNA as follows:

tcdb Gene Isolation	[Stock]	[Final]	1X Vol (µL)
Q5 RXN buffer	5X	1X	4
For Primer "TcdB_F"	10 µM	0.5 µM	1
Rev Primer "TcdB_R"	10 µM	0.5 µM	1
C. diff genomic DNA	20 ng/µL	20 ng	1
dNTP's	10 mM	0.3 mM	0.6
Q5 High Fidelity DNA polymerase	2 U/µL	0.004 U	0.2
Water	-	-	12.2
Total (µL)	-	-	20

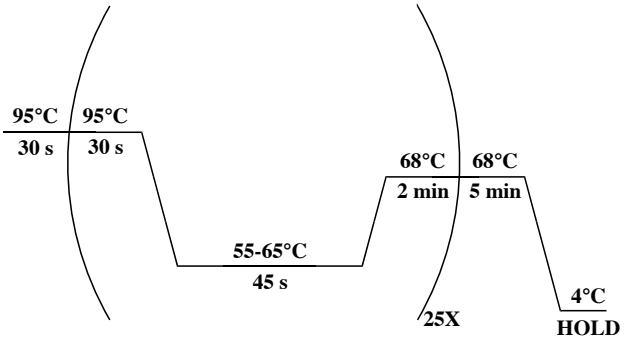


Primers 1 and 2 were equipped with BsrG1 (forward) and Kpn1 (reverse) cut sites and an annealing temperature of 53.4 °C was used to move forward. The *tcdb* amplicon was purified using MicroElute Cycle Pure Kit Centrifugation Protocol (OMEGA), eluted in 10 mM Tris-HCl pH 8.5, and quantified by A_{260}. pHis1522 (MoBiTec) and the *tcdb* amplicon then underwent double digestion with BsrG1 and Kpn1 restriction enzymes (3 µg DNA, 1X NEB 2.1 Reaction Buffer, 30 U BsrG1, and 30 U Kpn1) at 37 °C for 4 hours (0.2 U of calf intestinal alkaline phosphatase (CIP) was added to the pHis1522 plasmid digest reaction at the 3 hour mark). Next, reactions were mixed with 1X crystal violet loading dye (5% glycerol, 3.3 mM EDTA, 16.6 µg/mL crystal violet), loaded onto a 0.8% agarose gel containing 1.6 mg/mL crystal violet, separated by gel electrophoresis (80 V, 1 hr), and purified using the EZNA Gel Extraction Kit (OMEGA). After quantitation at A_{260}, pHis1522 and *tcdb* insert were ligated together (20 U T4

DNA ligase, 1X T4 buffer, 50 ng pHis1522, 37.5 ng *tcdb* DNA) at RT for 15 min following a 10 min heat-inactivation period at 65 °C, transforming into top10 *E. coli* cells by electroporation (1.5 µL ligation rxn, 75 µL competent cells), outgrowing in 200 µL standard LB broth (37 °C, 250 rpm, 30 min), plating on 100 µg/mL ampicillin plates, and incubating at 37 °C overnight. Resulting colonies were miniprepped (OMEGA EZNA Plasmid DNA Mini Kit II) and sequenced (Wayne State University School of Medicine Applied Genomics Technology Center) with primers seen in **Table 3**. Successful sequencing of *rtcdb* is shown in **Appendix I**. pHis1522-*tcdb* (pBMJ_130401) was transformed into *B. meg,* expressed, and purified as stated previously (**Chapter 2 Methods**).

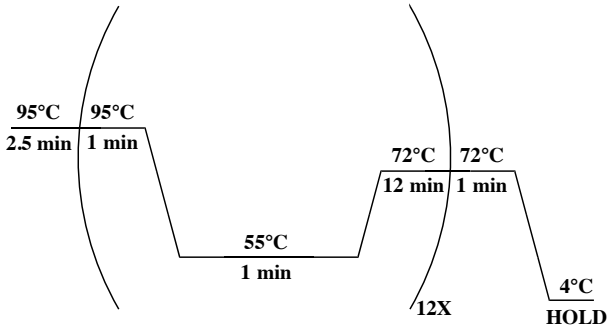
Clostridium difficile rTcdB-NVN cloning. Primers 1 and 3 from **Table 2** were used to isolate and amplify *rtcdb*-GTD as follows:

tcdb-GTD Isolation	[Stock]	[Final]	1X Vol (µL)
taq RXN buffer	10X	1X	2
For Primer "TcdB_F"	10 µM	0.5 µM	1
Rev Primer "B2R_2"	10 µM	0.5 µM	1
DNA (pBMJ_130401)	2.86 ng/µL	0.5 ng	0.175
$MgCl_2$	15 mM	1.5 mM	2
dNTP's	10 mM	0.5 mM	1
taq polymerase	5 U/µL	0.625 U	0.125
Water	-	-	12.7
Total (µL)	-	-	20



The PCR reaction with annealing temperature of 61.1 °C was used to move forward. Ligation of *tcdb*-GTD into pCR-XL-TOPO vector was performed at a 3:1 insert to vector ratio (RT, 30 min) before transforming into top 10 *E. coli* cells and plating on amp LB-agar plates as above. Selected colonies were then miniprepped (OMEGA EZNA Plasmid DNA Mini Kit II) and resulting DNA (pBMJ_70401) was used with primers 4 and 5 (**Table 2**) to create D286N/D288N mutations as follows:

tcdb-NVN SDM PCR	[Stock]	[Final]	1X Vol (µL)
***pfu* RXN buffer**	10X	1X	2.5
For Primer "rTcdB_NVN_F2"	10 µM	0.25 µM	0.625
Rev Primer "rTcdB_NVN_R2"	10 µM	0.25 µM	0.625
DNA (pBMJ_70401)	10 ng/µL	10 ng	1
dNTP's	10 mM	0.2 mM	0.5
***pfu* DNA polymerase**	5 U/µL	1.25 U	0.25
Water	-	-	19.5
Total (µL)	-	-	25

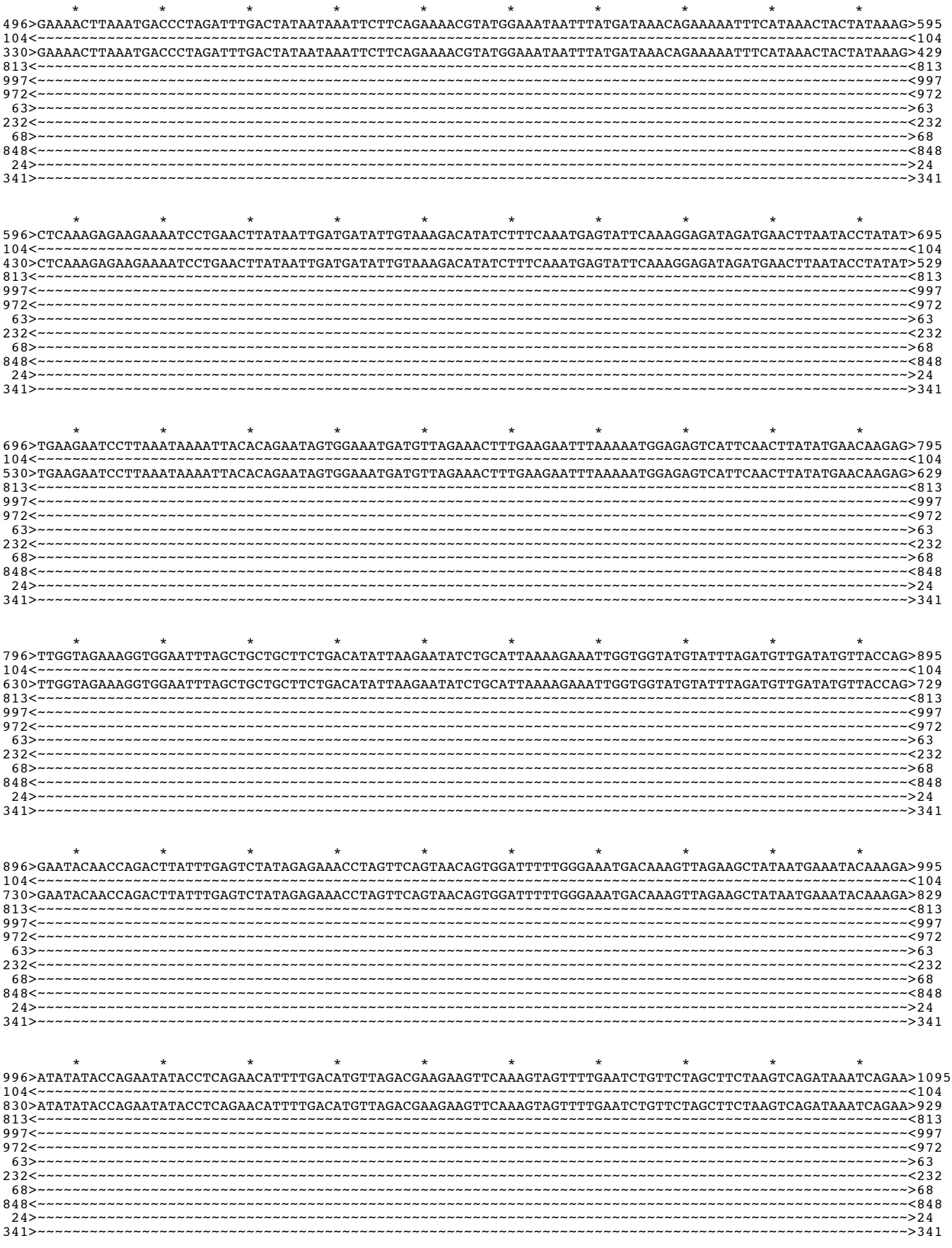


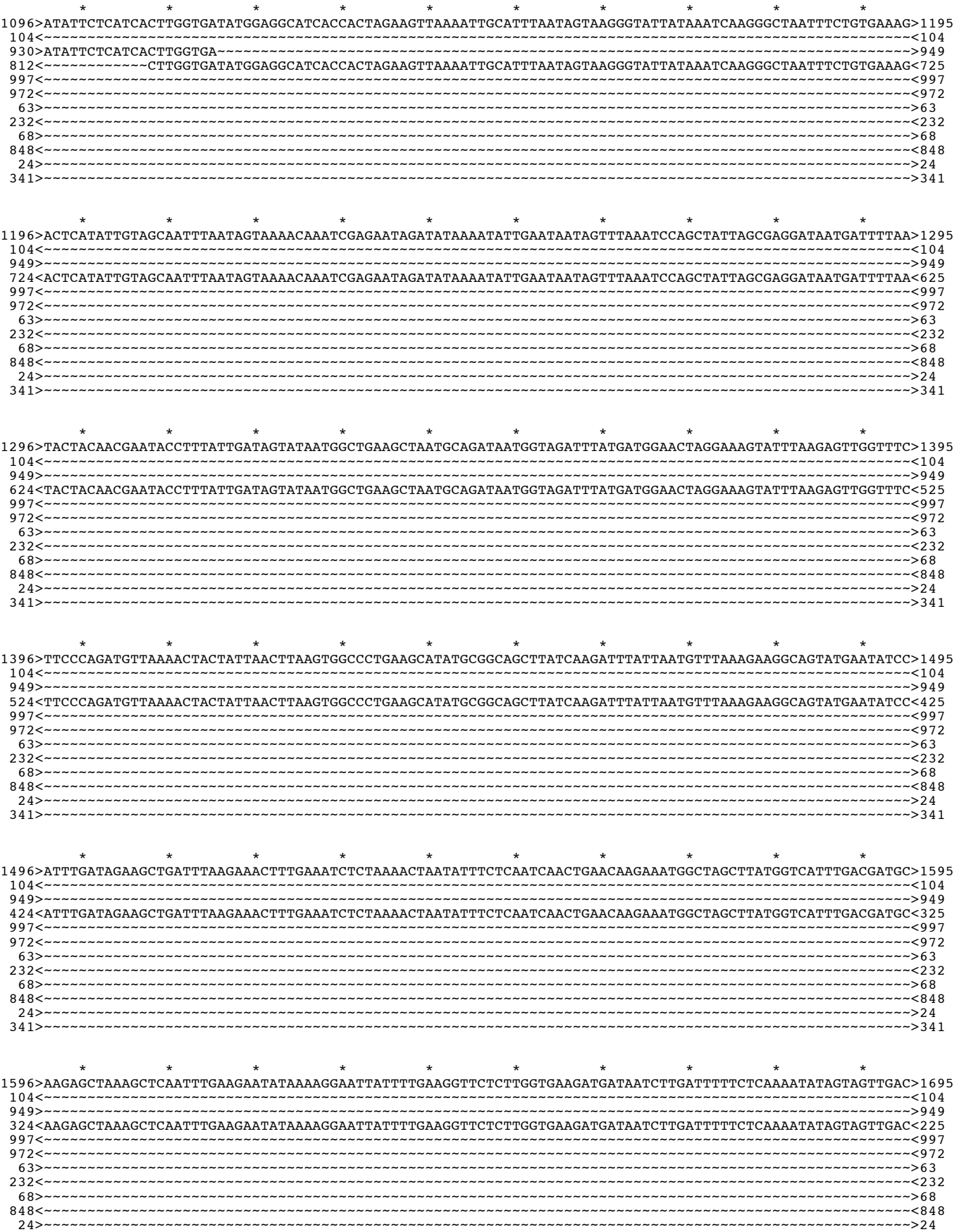
Sequencing of pBMJ_70402 was done using primers in **Table 4** and verification can be seen in **Appendix II**.

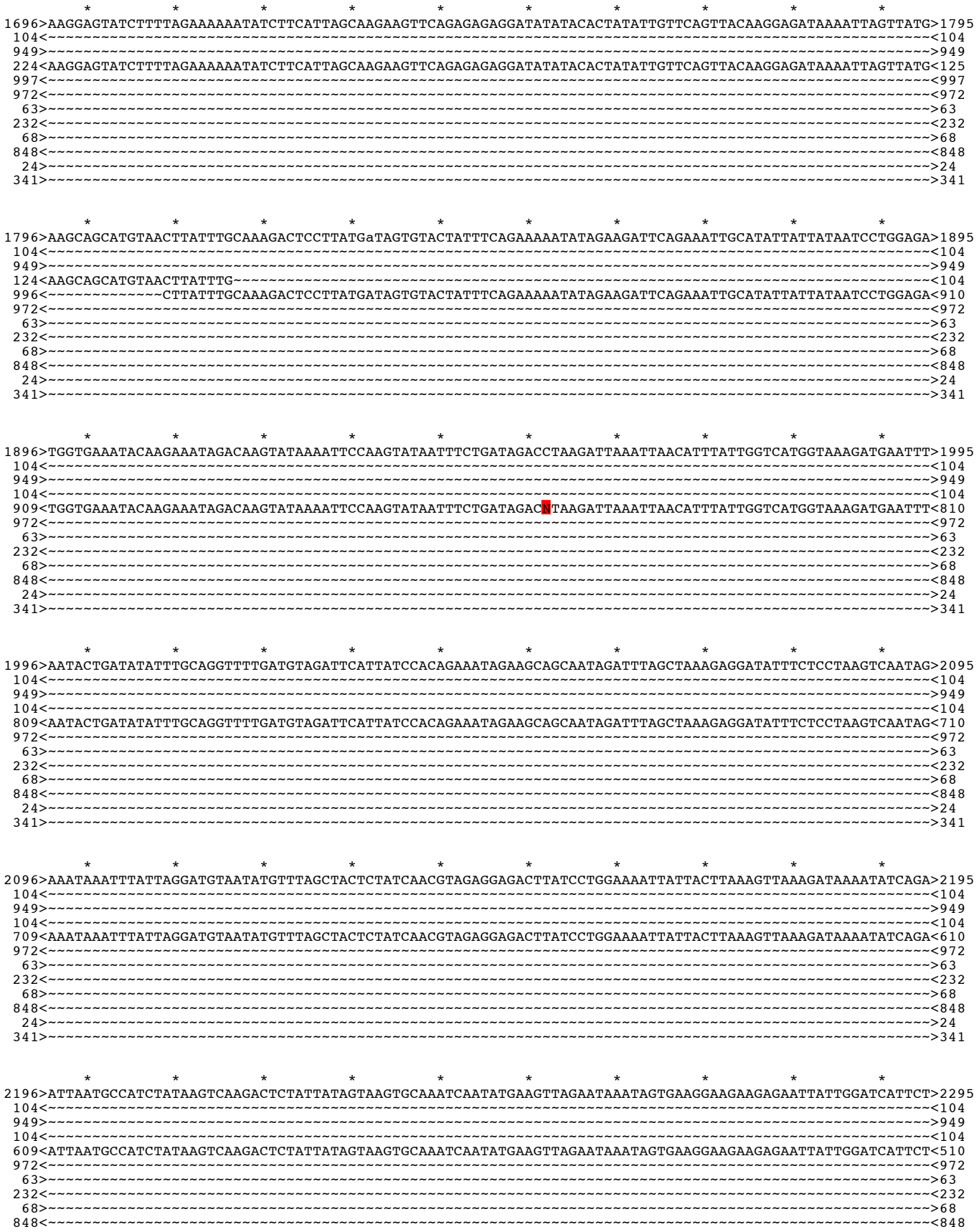
HeLa cell cultures. HeLa cell cultures were obtained from the Pflum Lab (Wayne State University) and contained to BL2 standard protocols. See **Chapter 3 Methods** for further maintenance procedures and protocols utilizing HeLa cells. Visual cell killing assays were done as described in **Chapter 2 Methods** with the exception of adding HQS-Epoxy inhibitor peptide.

APPENDIX I: rTcdB-WT Sequence Alignment

```
FINAL_pHIS1522_TcdB_BMJ.ape (pBMJ_130401; colony 1). Any errors/mismatches were confirmed manually by chromatogram.
From 16 to 7126 alignment to
3591_B01_TcdB1__B1R_007.seq--  Matches:225; Mismatches:0; Gaps:6906; Unattempted:0
3591_C01_TcdB1__B2F_006.seq--  Matches:882; Mismatches:0; Gaps:6229; Unattempted:0
3591_D01_TcdB1__B2R_005.seq--  Matches:709; Mismatches:0; Gaps:6402; Unattempted:0
3591_F01_TcdB1__B3R_003.seq--  Matches:849; Mismatches:1; Gaps:6261; Unattempted:0
3591_H01_TcdB1__B4R_001.seq--  Matches:766; Mismatches:0; Gaps:6345; Unattempted:0
3591_A02_TcdB1__B5F_008.seq--  Matches:816; Mismatches:0; Gaps:6295; Unattempted:0
3591_B02_TcdB1__B5R_007.seq--  Matches:197; Mismatches:2; Gaps:6912; Unattempted:0
3591_C02_TcdB1__B6F_006.seq--  Matches:444; Mismatches:0; Gaps:6667; Unattempted:0
3591_D02_TcdB1__B6R_005.seq--  Matches:752; Mismatches:0; Gaps:6359; Unattempted:0
3591_E02_TcdB1__B7F_004.seq--  Matches:995; Mismatches:0; Gaps:6116; Unattempted:0
3591_G02_TcdB1__B8F_002.seq--  Matches:579; Mismatches:0; Gaps:6543; Unattempted:0

         *         *         *         *         *         *         *         *         *         *
 16>~~~~~~~~~~~~~~~~~tgtacaATGAGTTTAGTTAATAGAAAACAGTTAGAAAAAATGGCAAATGTAAGATTTCGTACTCAAGAAGATGAATATGT>95
348<ACTTAAATCAAAGGGGGAAATGTACAATGAGTTTAGTTAATAGAAAACAGTTAGAAAAAATGGCAAATGTAAGATTTCGTACTCAAGAAGATGAATATGT<249
 67>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>67
813<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<813
997<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<997
972<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<972
 63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
 68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
 24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


         *         *         *         *         *         *         *         *         *         *
 96>TGCAATATTGGATGCTTTAGAAGAATATCATAATATGTCAGAGAATACTGTAGTCGAAAAATATTTAAAATTAAAAGATATAAATAGTTTAACAGATATT>195
248<TGCAATATTGGATGCTTTAGAAGAATATCATAATATGTCAGAGAATACTGTAGTCGAAAAATATTTAAAATTAAAAGATATAAATAGTTTAACAGATATT<149
 67>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>67
813<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<813
997<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<997
972<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<972
 63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
 68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
 24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


         *         *         *         *         *         *         *         *         *         *
196>TATATAGATACATATAAAAAATCTGGTAGAAATAAAGCCTTAAAAAAATTTAAGGAATATCTAGTTACAGAAGTATTAGAGCTAAAGAATAATAATTTAA>295
148<TATATAGATACATATAAAAAATCTGGTAGAAATAAAGCCTTAAAA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CTTAAAAAAATTTAAGGAATATCTAGTTACAGAAGTATTAGAGCTAAAGAATAATAATTTAA>129
813<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<813
997<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<997
972<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<972
 63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
 68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
 24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


         *         *         *         *         *         *         *         *         *         *
296>CTCCAGTTGAGAAAAATTTACATTTTGTTTGGATTGGAGGTCAAATAAATGACACTGCTATTAATTATATAAATCAATGGAAAGATGTAAATAGTGATTA>395
104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
130>CTCCAGTTGAGAAAAATTTACATTTTGTTTGGATTGGAGGTCAAATAAATGACACTGCTATTAATTATATAAATCAATGGAAAGATGTAAATAGTGATTA>229
813<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<813
997<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<997
972<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<972
 63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
 68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
 24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


         *         *         *         *         *         *         *         *         *         *
396>TAATGTTAATGTTTTTTATGATAGTAATGCATTTTTGATAAACACATTGAAAAAAACTGTAGTAGAATCAGCAATAAATGATACACTTGAATCATTTAGA>495
104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
230>TAATGTTAATGTTTTTTATGATAGTAATGCATTTTTGATAAACACATTGAAAAAAACTGTAGTAGAATCAGCAATAAATGATACACTTGAATCATTTAGA>329
813<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<813
997<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<997
972<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<972
 63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
 68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
 24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341
```


496>GAAAACTTAAATGACCCTAGATTTGACTATAATAAATTCTTCAGAAAACGTATGGAAATAATTTATGATAAACAGAAAAATTTCATAAACTACTATAAAG>595
104<104
330>GAAAACTTAAATGACCCTAGATTTGACTATAATAAATTCTTCAGAAAACGTATGGAAATAATTTATGATAAACAGAAAAATTTCATAAACTACTATAAAG>429
813<813
997<997
972<972
63>63
232<232
68>68
848<848
24>24
341>341
596>CTCAAAGAGAAGAAAATCCTGAACTTATAATTGATGATATTGTAAAGACATATCTTTCAAATGAGTATTCAAAGGAGATAGATGAACTTAATACCTATAT>695
104<104
430>CTCAAAGAGAAGAAAATCCTGAACTTATAATTGATGATATTGTAAAGACATATCTTTCAAATGAGTATTCAAAGGAGATAGATGAACTTAATACCTATAT>529
813<813
997<997
972<972
63>63
232<232
68>68
848<848
24>24
341>341
696>TGAAGAATCCTTAAATAAAATTACACAGAATAGTGGAAATGATGTTAGAAACTTTGAAGAATTTAAAAATGGAGAGTCATTCAACTTATATGAACAAGAG>795
104<104
530>TGAAGAATCCTTAAATAAAATTACACAGAATAGTGGAAATGATGTTAGAAACTTTGAAGAATTTAAAAATGGAGAGTCATTCAACTTATATGAACAAGAG>629
813<813
997<997
972<972
63>63
232<232
68>68
848<848
24>24
341>341
796>TTGGTAGAAAGGTGGAATTTAGCTGCTGCTTCTGACATATTAAGAATATCTGCATTAAAAGAAATTGGTGGTATGTATTTAGATGTTGATATGTTACCAG>895
104<104
630>TTGGTAGAAAGGTGGAATTTAGCTGCTGCTTCTGACATATTAAGAATATCTGCATTAAAAGAAATTGGTGGTATGTATTTAGATGTTGATATGTTACCAG>729
813<813
997<997
972<972
63>63
232<232
68>68
848<848
24>24
341>341
896>GAATACAACCAGACTTATTTGAGTCTATAGAGAAACCTAGTTCAGTAACAGTGGATTTTTGGGAAATGACAAAGTTAGAAGCTATAATGAAATACAAAGA>995
104<104
730>GAATACAACCAGACTTATTTGAGTCTATAGAGAAACCTAGTTCAGTAACAGTGGATTTTTGGGAAATGACAAAGTTAGAAGCTATAATGAAATACAAAGA>829
813<813
997<997
972<972
63>63
232<232
68>68
848<848
24>24
341>341
996>ATATATACCAGAATATACCTCAGAACATTTTGACATGTTAGACGAAGAAGTTCAAAGTAGTTTTGAATCTGTTCTAGCTTCTAAGTCAGATAAATCAGAA>1095
104<104
830>ATATATACCAGAATATACCTCAGAACATTTTGACATGTTAGACGAAGAAGTTCAAAGTAGTTTTGAATCTGTTCTAGCTTCTAAGTCAGATAAATCAGAA>929
813<813
997<997
972<972
63>63
232<232
68>68
848<848
24>24
341>341

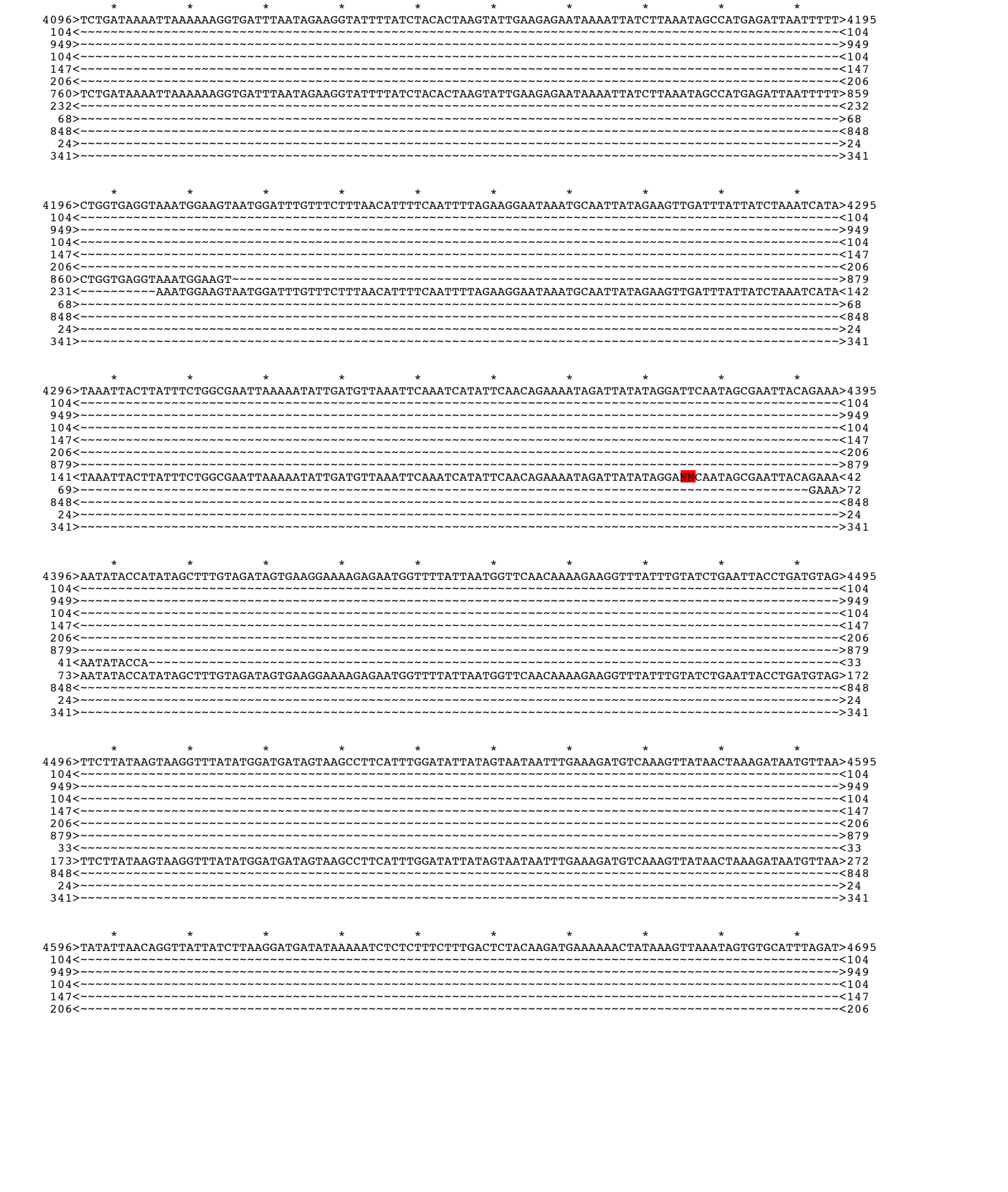

1096>ATATTCTCATCACTTGGTGATATGGAGGCATCACCACTAGAAGTTAAAATTGCATTTAATAGTAAGGGTATTATAAATCAAGGGCTAATTTCTGTGAAAG>1195
104<~~~~~~<104
930>ATATTCTCATCACTTGGTGA~~~~~~>949
812<~~~~~~~~~~~~CTTGGTGATATGGAGGCATCACCACTAGAAGTTAAAATTGCATTTAATAGTAAGGGTATTATAAATCAAGGGCTAATTTCTGTGAAAG<725
997<~~~~~~<997
972<~~~~~~<972
63>~~~~~~>63
232<~~~~~~<232
68>~~~~~~>68
848<~~~~~~<848
24>~~~~~~>24
341>~~~~~~>341
1196>ACTCATATTGTAGCAATTTAATAGTAAAACAAATCGAGAATAGATATAAAATATTGAATAATAGTTTAAATCCAGCTATTAGCGAGGATAATGATTTTAA>1295
104<~~~~~~<104
949>~~~~~~>949
724<ACTCATATTGTAGCAATTTAATAGTAAAACAAATCGAGAATAGATATAAAATATTGAATAATAGTTTAAATCCAGCTATTAGCGAGGATAATGATTTTAA<625
997<~~~~~~<997
972<~~~~~~<972
63>~~~~~~>63
232<~~~~~~<232
68>~~~~~~>68
848<~~~~~~<848
24>~~~~~~>24
341>~~~~~~>341
1296>TACTACAACGAATACCTTTATTGATAGTATAATGGCTGAAGCTAATGCAGATAATGGTAGATTTATGATGGAACTAGGAAAGTATTTAAGAGTTGGTTTC>1395
104<~~~~~~<104
949>~~~~~~>949
624<TACTACAACGAATACCTTTATTGATAGTATAATGGCTGAAGCTAATGCAGATAATGGTAGATTTATGATGGAACTAGGAAAGTATTTAAGAGTTGGTTTC<525
997<~~~~~~<997
972<~~~~~~<972
63>~~~~~~>63
232<~~~~~~<232
68>~~~~~~>68
848<~~~~~~<848
24>~~~~~~>24
341>~~~~~~>341
1396>TTCCCAGATGTTAAAACTACTATTAACTTAAGTGGCCCTGAAGCATATGCGGCAGCTTATCAAGATTTATTAATGTTTAAAGAAGGCAGTATGAATATCC>1495
104<~~~~~~<104
949>~~~~~~>949
524<TTCCCAGATGTTAAAACTACTATTAACTTAAGTGGCCCTGAAGCATATGCGGCAGCTTATCAAGATTTATTAATGTTTAAAGAAGGCAGTATGAATATCC<425
997<~~~~~~<997
972<~~~~~~<972
63>~~~~~~>63
232<~~~~~~<232
68>~~~~~~>68
848<~~~~~~<848
24>~~~~~~>24
341>~~~~~~>341
1496>ATTTGATAGAAGCTGATTTAAGAAACTTTGAAATCTCTAAAACTAATATTTCTCAATCAACTGAACAAGAAATGGCTAGCTTATGGTCATTTGACGATGC>1595
104<~~~~~~<104
949>~~~~~~>949
424<ATTTGATAGAAGCTGATTTAAGAAACTTTGAAATCTCTAAAACTAATATTTCTCAATCAACTGAACAAGAAATGGCTAGCTTATGGTCATTTGACGATGC<325
997<~~~~~~<997
972<~~~~~~<972
63>~~~~~~>63
232<~~~~~~<232
68>~~~~~~>68
848<~~~~~~<848
24>~~~~~~>24
341>~~~~~~>341
1596>AAGAGCTAAAGCTCAATTTGAAGAATATAAAAGGAATTATTTTGAAGGTTCTCTTGGTGAAGATGATAATCTTGATTTTTCTCAAAATATAGTAGTTGAC>1695
104<~~~~~~<104
949>~~~~~~>949
324<AAGAGCTAAAGCTCAATTTGAAGAATATAAAAGGAATTATTTTGAAGGTTCTCTTGGTGAAGATGATAATCTTGATTTTTCTCAAAATATAGTAGTTGAC<225
997<~~~~~~<997
972<~~~~~~<972
63>~~~~~~>63
232<~~~~~~<232
68>~~~~~~>68
848<~~~~~~<848
24>~~~~~~>24


1696>AAGGAGTATCTTTTAGAAAAAATATCTTCATTAGCAAGAAGTTCAGAGAGAGGATATATACACTATATTGTTCAGTTACAAGGAGATAAAATTAGTTATG>1795
104<104
949>949
224<AAGGAGTATCTTTTAGAAAAAATATCTTCATTAGCAAGAAGTTCAGAGAGAGGATATATACACTATATTGTTCAGTTACAAGGAGATAAAATTAGTTATG<125
997<997
972<972
63>63
232<232
68>68
848<848
24>24
341>341
1796>AAGCAGCATGTAACTTATTTGCAAAGACTCCTTATGaTAGTGTACTATTTCAGAAAAATATAGAAGATTCAGAAATTGCATATTATTATAATCCTGGAGA>1895
104<104
949>949
124<AAGCAGCATGTAACTTATTTG
996<CTTATTTGCAAAGACTCCTTATGATAGTGTACTATTTCAGAAAAATATAGAAGATTCAGAAATTGCATATTATTATAATCCTGGAGA<910
972<972
63>63
232<232
68>68
848<848
24>24
341>341
1896>TGGTGAAATACAAGAAATAGACAAGTATAAAATTCCAAGTATAATTTCTGATAGACCTAAGATTAAATTAACATTTATTGGTCATGGTAAAGATGAATTT>1995
104<104
949>949
104<104
909<TGGTGAAATACAAGAAATAGACAAGTATAAAATTCCAAGTATAATTTCTGATAGACNTAAGATTAAATTAACATTTATTGGTCATGGTAAAGATGAATTT<810
972<972
63>63
232<232
68>68
848<848
24>24
341>341
1996>AATACTGATATATTTGCAGGTTTTGATGTAGATTCATTATCCACAGAAATAGAAGCAGCAATAGATTTAGCTAAAGAGGATATTTCTCCTAAGTCAATAG>2095
104<104
949>949
104<104
809<AATACTGATATATTTGCAGGTTTTGATGTAGATTCATTATCCACAGAAATAGAAGCAGCAATAGATTTAGCTAAAGAGGATATTTCTCCTAAGTCAATAG<710
972<972
63>63
232<232
68>68
848<848
24>24
341>341
2096>AAATAAATTTATTAGGATGTAATATGTTTAGCTACTCTATCAACGTAGAGGAGACTTATCCTGGAAAATTATTACTTAAAGTTAAAGATAAAATATCAGA>2195
104<104
949>949
104<104
709<AAATAAATTTATTAGGATGTAATATGTTTAGCTACTCTATCAACGTAGAGGAGACTTATCCTGGAAAATTATTACTTAAAGTTAAAGATAAAATATCAGA<610
972<972
63>63
232<232
68>68
848<848
24>24
341>341
2196>ATTAATGCCATCTATAAGTCAAGACTCTATTATAGTAAGTGCAAATCAATATGAAGTTAGAATAAATAGTGAAGGAAGAAGAGAATTATTGGATCATTCT>2295
104<104
949>949
104<104
609<ATTAATGCCATCTATAAGTCAAGACTCTATTATAGTAAGTGCAAATCAATATGAAGTTAGAATAAATAGTGAAGGAAGAAGAGAATTATTGGATCATTCT<510
972<972
63>63
232<232
68>68
848<848

```
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


         *         *         *         *         *         *         *         *         *         *
2296>GGTGAATGGATAAATAAAGAAGAAAGTATTATAAAGGATATTTCATCAAAAGAATATATATCATTTAATCCTAAAGAAAATAAAATTACAGTAAAATCTA>2395
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 509<GGTGAATGGATAAATAAAGAAGAAAGTATTATAAAGGATATTTCATCAAAAGAATATATATCATTTAATCCTAAAGAAAATAAAATTACAGTAAAATCTA<410
 972<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<972
  63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


         *         *         *         *         *         *         *         *         *         *
2396>AAAATTTACCTGAGCTATCTACATTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAAAAAGTAATGTTAACAGAATGTGA>2495
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 409<AAAATTTACCTGAGCTATCTACATTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAAAAAGTAATGTTAACAGAATGTGA<310
 972<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<972
  63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


         *         *         *         *         *         *         *         *         *         *
2496>GATAAATGTTATTTCAAATATAGATACGCAAATTGTTGAGGAAAGGATTGAAGAAGCTAAGAATTTAACTTCTGACTCTATTAATTATATAAAAGATGAA>2595
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 309<GATAAATGTTATTTCAAATATAGATACGCAAATTGTTGAGGAAAGGATTGAAGAAGCTAAGAATTTAACTTCTGACTCTATTAATTATATAAAAGATGAA<210
 972<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<972
  63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


         *         *         *         *         *         *         *         *         *         *
2596>TTTAAACTAATAGAATCTATTTCTGATGCACTATGTGACTTAAAACAACAGAATGAATTAGAAGATTCTCATTTTATATCTTTTGAGGACATATCAGAGA>2695
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 209<TTTAAACTAATAGAATCTATTTCTGATGCACTATGTGACTTAAAACAACAGAATGAATTAGAA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 971<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAATGAATTAGAAGATTCTCATTTTATATCTTTTGAGGACATATCAGAGA<922
  63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


         *         *         *         *         *         *         *         *         *         *
2696>CTGATGAGGGATTTAGTATAAGATTTATTAATAAAGAAACTGGAGAATCTATATTTGTAGAAACTGAAAAAACAATATTCTCTGAATATGCTAATCATAT>2795
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 921<CTGATGAGGGATTTAGTATAAGATTTATTAATAAAGAAACTGGAGAATCTATATTTGTAGAAACTGAAAAAACAATATTCTCTGAATATGCTAATCATAT<822
  63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


         *         *         *         *         *         *         *         *         *         *
2796>AACTGAAGAGATTTCTAAGATAAAAGGTACTATATTTGATACTGTAAATGGTAAGTTAGTAAAAAAAGTAAATTTAGATACTACACACGAAGTAAATACT>2895
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 821<AACTGAAGAGATTTCTAAGATAAAAGGTACTATATTTGATACTGTAAATGGTAAGTTAGTAAAAAAAGTAAATTTAGATACTACACACGAAGTAAATACT<722
  63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
```

```
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
2896>TTAAATGCTGCATTTTTTATACAATCATTAATAGAATATAATAGTTCTAAAGAATCTCTTAGTAATTTAAGTGTAGCAATGAAAGTCCAAGTTTACGCTC>2995
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 721<TTAAATGCTGCATTTTTTATACAATCATTAATAGAATATAATAGTTCTAAAGAATCTCTTAGTAATTTAAGTGTAGCAATGAAAGTCCAAGTTTACGCTC<622
  63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
2996>AATTATTTAGTACTGGTTTAAATACTATTACAGATGCAGCCAAAGTTGTTGAATTAGTATCAACTGCATTAGATGAAACTATAGACTTACTTCCTACATT>3095
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 621<AATTATTTAGTACTGGTTTAAATACTATTACAGATGCAGCCAAAGTTGTTGAATTAGTATCAACTGCATTAGATGAAACTATAGACTTACTTCCTACATT<522
  63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
3096>ATCTGAAGGATTACCTATAATTGCAACTATTATAGATGGTGTAAGTTTAGGTGCAGCAATCAAAGAGCTAAGTGAAACGAGTGACCCATTATTAAGACAA>3195
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 521<ATCTGAAGGATTACCTATAATTGCAACTATTATAGATGGTGTAAGTTTAGGTGCAGCAATCAAAGAGCTAAGTGAAACGAGTGACCCATTATTAAGACAA<422
  63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
3196>GAAATAGAAGCTAAGATAGGTATAATGGCAGTAAATTTAACAACAGCTACAACTGCAATCATTACTTCATCTTTGGGGgTAGCTAGTGGATTTAGTATAC>3295
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 421<GAAATAGAAGCTAAGATAGGTATAATGGCAGTAAATTTAACAACAGCTACAACTGCAATCATTACTTCATCTTTGGGGGTAGCTAGTGGATTTAGTATAC<322
  63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
3296>TTTTAGTTCCTTTAGCAGGAATTTCAGCAGGTATACCAAGCTTAGTAAACAATGAACTTGTACTTCGAGATAAGGCAACAAAGGTTGTAGATTATTTTAA>3395
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 321<TTTTAGTTCCTTTAGCAGGAATTTCAGCAGGTATACCAAGCTTAGTAAACAATGAACTTGTACTTCGAGATAAGGCAACAAAGGTTGTAGATTATTTTAA<222
  63>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>63
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
3396>ACATGTTTCATTAGTTGAAACTGAAGGAGTATTTACTTTATTAGATGATAAAATAATGATGCCACAAGATGATTTAGTGATATCAGAAATAGATTTTAAT>3495
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 221<ACATGTTTCATTAGTT~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
  64>~~~~GTTTCATTAGTTGAAACTGAAGGAGTATTTACTTTATTAGATGATAAAATAATGATGCCACAAGATGATTTAGTGATATCAGAAATAGATTTTAAT>159
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
```

```
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

              *         *         *         *         *         *         *         *         *         *
3496>AATAATTCAATAGTTTTAGGTAAATGTGAAATCTGGAGAATGGAAGGTGGTTCAGGTCATACTGTAACTGATGATATAGATCACTTCTTTTCAGCACCAT>3595
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 160>AATAATTCAATAGTTTTAGGTAAATGTGAAATCTGGAGAATGGAAGGTGGTTCAGGTCATACTGTAACTGATGATATAGATCACTTCTTTTCAGCACCAT>259
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

              *         *         *         *         *         *         *         *         *         *
3596>CAATAACATATAGAGAGCCACACTTATCTATATATGACGTATTGGAAGTACAAAAAGAAGAACTTGATTTGTCAAAAGATTTAATGGTATTACCTAATGC>3695
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 260>CAATAACATATAGAGAGCCACACTTATCTATATATGACGTATTGGAAGTACAAAAAGAAGAACTTGATTTGTCAAAAGATTTAATGGTATTACCTAATGC>359
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

              *         *         *         *         *         *         *         *         *         *
3696>TCCAAATAGAGTATTTGCTTGGGAAACAGGATGGACACCAGGTTTAAGAAGCTTAGAAAATGATGGCACAAAACTGTTAGACCGTATAAGAGATAACTAT>3795
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 360>TCCAAATAGAGTATTTGCTTGGGAAACAGGATGGACACCAGGTTTAAGAAGCTTAGAAAATGATGGCACAAAACTGTTAGACCGTATAAGAGATAACTAT>459
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

              *         *         *         *         *         *         *         *         *         *
3796>GAAGGTGAGTTTTATTGGAGATATTTTGCTTTTATAGCTGATGCTTTAATAACAACATTAAAACCAAGATATGAAGATACTAATATAAGAATAAATTTAG>3895
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 460>GAAGGTGAGTTTTATTGGAGATATTTTGCTTTTATAGCTGATGCTTTAATAACAACATTAAAACCAAGATATGAAGATACTAATATAAGAATAAATTTAG>559
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

              *         *         *         *         *         *         *         *         *         *
3896>ATAGTAATACTAGAAGTTTTATAGTTCCAATAATAACTACAGAATATATAAGAGAAAAATTATCATATTCTTTCTATGGTTCAGGAGGAACTTATGCATT>3995
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 560>ATAGTAATACTAGAAGTTTTATAGTTCCAATAATAACTACAGAATATATAAGAGAAAAATTATCATATTCTTTCTATGGTTCAGGAGGAACTTATGCATT>659
 232<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<232
  68>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>68
 848<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<848
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

              *         *         *         *         *         *         *         *         *         *
3996>GTCTCTTTCTCAATATAATATGGGTATAAATATAGAATTAAGTGAAAGTGATGTTTGGATTATAGATGTTGATAATGTTGTGAGAGATGTAACTATAGAA>4095
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 660>GTCTCTTTCTCAATATAATATGGGTATAAATATAGAATTAAGTGAAAGTGATGTTTGGATTATAGATGTTGATAATGTTGTGAGAGATGTAACTATAGAA>759
```


4096>TCTGATAAAATTAAAAAAGGTGATTTAATAGAAGGTATTTTATCTACACTAAGTATTGAAGAGAATAAAATTATCTTAAATAGCCATGAGATTAATTTTT>4195
104<104
949>949
104<104
147<147
206<206
760>TCTGATAAAATTAAAAAAGGTGATTTAATAGAAGGTATTTTATCTACACTAAGTATTGAAGAGAATAAAATTATCTTAAATAGCCATGAGATTAATTTTT>859
232<232
68>68
848<848
24>24
341>341
4196>CTGGTGAGGTAAATGGAAGTAATGGATTTGTTTCTTTAACATTTTCAATTTTAGAAGGAATAAATGCAATTATAGAAGTTGATTTATTATCTAAATCATA>4295
104<104
949>949
104<104
147<147
206<206
860>CTGGTGAGGTAAATGGAAGT~>879
231<~~~~~~~~~~AAATGGAAGTAATGGATTTGTTTCTTTAACATTTTCAATTTTAGAAGGAATAAATGCAATTATAGAAGTTGATTTATTATCTAAATCATA<142
68>68
848<848
24>24
341>341
4296>TAAATTACTTATTTCTGGCGAATTAAAAATATTGATGTTAAATTCAAATCATATTCAACAGAAAATAGATTATATAGGATTCAATAGCGAATTACAGAAA>4395
104<104
949>949
104<104
147<147
206<206
879>879
141<TAAATTACTTATTTCTGGCGAATTAAAAATATTGATGTTAAATTCAAATCATATTCAACAGAAAATAGATTATATAGGANNCAATAGCGAATTACAGAAA<42
69>~~~~~~GAAA>72
848<848
24>24
341>341
4396>AATATACCATATAGCTTTGTAGATAGTGAAGGAAAAGAGAATGGTTTTATTAATGGTTCAACAAAAGAAGGTTTATTTGTATCTGAATTACCTGATGTAG>4495
104<104
949>949
104<104
147<147
206<206
879>879
41<AATATACCA~>33
73>AATATACCATATAGCTTTGTAGATAGTGAAGGAAAAGAGAATGGTTTTATTAATGGTTCAACAAAAGAAGGTTTATTTGTATCTGAATTACCTGATGTAG>172
848<848
24>24
341>341
4496>TTCTTATAAGTAAGGTTTATATGGATGATAGTAAGCCTTCATTTGGATATTATAGTAATAATTTGAAAGATGTCAAAGTTATAACTAAAGATAATGTTAA>4595
104<104
949>949
104<104
147<147
206<206
879>879
33<33
173>TTCTTATAAGTAAGGTTTATATGGATGATAGTAAGCCTTCATTTGGATATTATAGTAATAATTTGAAAGATGTCAAAGTTATAACTAAAGATAATGTTAA>272
848<848
24>24
341>341
4596>TATATTAACAGGTTATTATCTTAAGGATGATATAAAAATCTCTCTTTCTTTGACTCTACAAGATGAAAAAACTATAAAGTTAAATAGTGTGCATTTAGAT>4695
104<104
949>949
104<104
147<147
206<206

879>~~~~~~>879
33<~~~~~~<33
273>TATATTAACAGGTTATTATCTTAAGGATGATATAAAAATCTCTCTTTCTTTGACTCTACAAGATGAAAAAACTATAAAGTTAAATAGTGTGCATTTAGAT>372
848<~~~~~~<848
24>~~~~~~>24
341>~~~~~~>341

* * * * * * * * * *
4696>GAAAGTGGAGTAGCTGAGATTTTGAAGTTCATGAATAGAAAAGGTAATACAAATACTTCAGATTCTTTAATGAGCTTTTTAGAAAGTATGAATATAAAAA>4795
104<~~~~~~<104
949>~~~~~~>949
104<~~~~~~<104
147<~~~~~~<147
206<~~~~~~<206
879>~~~~~~>879
33<~~~~~~<33
373>GAAAGTGGAGTAGCTGAGATTTTGAAGTTCATGAATAGAAAAGGTAATACAAATACTTCAGATTCTTTAATGAGCTTTTTAGAAAGTATGAATATAAAAA>472
848<~~~~~~<848
24>~~~~~~>24
341>~~~~~~>341

* * * * * * * * * *
4796>GTATTTTCGTTAATTTCTTACAATCTAATATTAAGTTTATATTAGATGCTAATTTTATAATAAGTGGTACTACTTCTATTGGCCAATTTGAGTTTATTTG>4895
104<~~~~~~<104
949>~~~~~~>949
104<~~~~~~<104
147<~~~~~~<147
206<~~~~~~<206
879>~~~~~~>879
33<~~~~~~<33
473>GTATTTTCGTTAATTTCTTACAATCTAATATTAAGTTTAT~~~~~~>512
847<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~AAGTTTATATTAGATGCTAATTTTATAATAAGTGGTACTACTTCTATTGGCCAATTTGAGTTTATTTG<780
24>~~~~~~>24
341>~~~~~~>341

* * * * * * * * * *
4896>TGATGAAAATGATAATATACAACCATATTTCATTAAGTTTAATACACTAGAAACTAATTATACTTTATATGTAGGAAATAGACAAAATATGATAGTGGAA>4995
104<~~~~~~<104
949>~~~~~~>949
104<~~~~~~<104
147<~~~~~~<147
206<~~~~~~<206
879>~~~~~~>879
33<~~~~~~<33
512>~~~~~~>512
779<TGATGAAAATGATAATATACAACCATATTTCATTAAGTTTAATACACTAGAAACTAATTATACTTTATATGTAGGAAATAGACAAAATATGATAGTGGAA<680
24>~~~~~~>24
341>~~~~~~>341

* * * * * * * * * *
4996>CCAAATTATGATTTAGATGATTCTGGAGATATATCTTCAACTGTTATCAATTTCTCTCAAAAGTATCTTTATGGAATAGACAGTTGTGTTAATAAAGTTG>5095
104<~~~~~~<104
949>~~~~~~>949
104<~~~~~~<104
147<~~~~~~<147
206<~~~~~~<206
879>~~~~~~>879
33<~~~~~~<33
512>~~~~~~>512
679<CCAAATTATGATTTAGATGATTCTGGAGATATATCTTCAACTGTTATCAATTTCTCTCAAAAGTATCTTTATGGAATAGACAGTTGTGTTAATAAAGTTG<580
24>~~~~~~>24
341>~~~~~~>341

* * * * * * * * * *
5096>TAATTTCACCAAATATTTATACAGATGAAATAAATATAACGCCTGTATATGAAACAAATAATACTTATCCAGAAGTTATTGTATTAGATGCAAATTATAT>5195
104<~~~~~~<104
949>~~~~~~>949
104<~~~~~~<104
147<~~~~~~<147
206<~~~~~~<206
879>~~~~~~>879
33<~~~~~~<33
512>~~~~~~>512
579<TAATTTCACCAAATATTTATACAGATGAAATAAATATAACGCCTGTATATGAAACAAATAATACTTATCCAGAAGTTATTGTATTAGATGCAAATTATAT<480
24>~~~~~~>24
341>~~~~~~>341

* * * * * * * * * *
5196>AAATGAAAAAATAAATGTTAATATCAATGATCTATCTATACGATATGTATGGAGTAATGATGGTAATGATTTTATTCTTATGTCAACTAGTGAAGAAAAT>5295
104<~~~~~~<104
949>~~~~~~>949
104<~~~~~~<104
147<~~~~~~<147

```
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
 479<AAATGAAAAAATAAATGTTAATATCAATGATCTATCTATACGATATGTATGGAGTAATGATGGTAATGATTTTATTCTTATGTCAACTAGTGAAGAAAAT<380
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


              *         *         *         *         *         *         *         *         *         *
5296>AAGGTGTCACAAGTTAAAATAAGATTCGTTAATGTTTTTAAAGATAAGACTTTGGCAAATAAGCTATCTTTTAACTTTAGTGATAAACAAGATGTACCTG>5395
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
 379<AAGGTGTCACAAGTTAAAATAAGATTCGTTAATGTTTTTAAAGATAAGACTTTGGCAAATAAGCTATCTTTTAACTTTAGTGATAAACAAGATGTACCTG<280
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


              *         *         *         *         *         *         *         *         *         *
5396>TAAGTGAAATAATCTTATCATTTACACCTTCATATTATGAGGATGGATTGATTGGCTATGATTTGGGTCTAGTTTCTTTATATAATGAGAAATTTTATAT>5495
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
 279<TAAGTGAAATAATCTTATCATTTACACCTTCATATTATGAGGATGGATTGATTGGCTATGATTTGGGTCTAGTTTCTTTATATAATGAGAAATTTTATAT<180
  24>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>24
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


              *         *         *         *         *         *         *         *         *         *
5496>TAATAACTTTGGAATGATGGTATCTGGATTAATATATATTAATGATTCATTATATTATTTAAACCACCAGTAAATAATTTGATAACTGGATTTGTGACT>5595
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
 179<TAATAACTTTGGAATGATGGTATCTGGATTAATATATATTAATGATTCATTATATTATTTAAACCACCAGTAAATAATTTGAT~~~~~~~~~~~~~~~~<96
  25>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CCAGTAAATAATTTGATAACTGGATTTGTGACT>57
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


              *         *         *         *         *         *         *         *         *         *
5596>GTAGGCGATGATAAATACTACTTTAATCCAATTAATGGTGGAGCTGCTTCAATTGGAGAGACAATAATTGATGACAAAAATTATTATTTCAACCAAAGTG>5695
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
  58>GTAGGCGATGATAAATACTACTTTAATCCAATTAATGGTGGAGCTGCTTCAATTGGAGAGACAATAATTGATGACAAAAATTATTATTTCAACCAAAGTG>157
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


              *         *         *         *         *         *         *         *         *         *
5696>GAGTGTTACAAACAGGTGTATTTAGTACAGAAGATGGATTTAAATATTTTGCCCCAGCTAATACACTTGATGAAAACCTAGAAGGAGAAGCAATTGATTT>5795
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
 158>GAGTGTTACAAACAGGTGTATTTAGTACAGAAGATGGATTTAAATATTTTGCCCCAGCTAATACACTTGATGAAAACCTAGAAGGAGAAGCAATTGATTT>257
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341


              *         *         *         *         *         *         *         *         *         *
5796>TACTGGAAAATTAATTATTGACGAAAATATTTATTATTTTGATGATAATTATAGAGGAGCTGTAGAATGGAAAGAATTAGATGGTGAAATGCACTATTTT>5895
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
```

```
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
 258>TACTGGAAAATTAATTATTGACGAAAATATTTATTATTTTGATGATAATTATAGAGGAGCTGTAGAATGGAAAGAATTAGATGGTGAAATGCACTATTTT>357
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
5896>AGCCCAGAAACAGGTAAAGCTTTTAAAGGTCTAAATCAAATAGGTGATTATAAATACTATTTCAATTCTGATGGAGTTATGCAAAAAGGATTTGTTAGTA>5995
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
 358>AGCCCAGAAACAGGTAAAGCTTTTAAAGGTCTAAATCAAATAGGTGATTATAAATACTATTTCAATTCTGATGGAGTTATGCAAAAAGGATTTGTTAGTA>457
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
5996>TAAATGATAATAAACACTATTTTGATGATTCTGGTGTTATGAAAGTAGGTTACACTGAAATAGATGGCAAGCATTTCTACTTTGCTGAAAACGGAGAAAT>6095
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
 458>TAAATGATAATAAACACTATTTTGATGATTCTGGTGTTATGAAAGTAGGTTACACTGAAATAGATGGCAAGCATTTCTACTTTGCTGAAAACGGAGAAAT>557
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
6096>GCAAATAGGAGTATTTAATACAGAAGATGGATTTAAATATTTTGCTCATCATAATGAAGATTTAGGAAATGAAGAAGGTGAAGAAATCTCATATTCTGGT>6195
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
 558>GCAAATAGGAGTATTTAATACAGAAGATGGATTTAAATATTTTGCTCATCATAATGAAGATTTAGGAAATGAAGAAGGTGAAGAAATCTCATATTCTGGT>657
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
6196>ATATTAAATTTCAATAATAAAATTTACTATTTTGATGATTCATTTACAGCTGTAGTTGGATGGAAAGATTTAGAGGATGGTTCAAAGTATTATTTTGATG>6295
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
 658>ATATTAAATTTCAATAATAAAATTTACTATTTTGATGATTCATTTACAGCTGTAGTTGGATGGAAAGATTTAGAGGATGGTTCAAAGTATTATTTTGATG>757
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
6296>AAGATACAGCAGAAGCATATATAGGTTTGTCATTAATAAATGATGGTCAATATTATTTTAATGATGATGGAATTATGCAAGTTGGATTTGTCACTATAAA>6395
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
 758>AAGATACAGCAGAAGCATATATAGGTTTGTCATTAATAAATGATGGTCAATATTATTTTAATGATGATGGAATTATGCAAGTTGGATTTGTCACTATAAA>857
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
6396>TGATAAAGTCTTCTACTTCTCTGACTCTGGAATTATAGAATCTGGAGTACAAAACATAGATGACAATTATTTCTATATAGATGATAATGGTATAGTTCAA>6495
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
```

```
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
 858>TGATAAAGTCTTCTACTTCTCTGACTCTGGAATTATAGAATCTGGAGTACAAAACATAGATGACAATTATTTCTATATAGATGATAATGGTATAGTTCAA>957
 341>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>341

         *         *         *         *         *         *         *         *         *         *
6496>ATTGGTGTATTTGATACTTCAGATGGATATAAATATTTTGCACCTGCTAATACTGTAAATGATAATATTTACGGACAAGCAGTTGAATATAGTGGTTTAG>6595
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
 958>ATTGGTGTATTTGATACTTCAGATGGATATAAATATTTTGCACCTGCTAATACTGTAAATGA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>1019
 342>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CTGTAAATGATAATATTTACGGACAAGCAGTTGAATATAGTGGTTTAG>389

         *         *         *         *         *         *         *         *         *         *
6596>TTAGAGTTGGGGAAGATGTATATTATTTTGGAGAAACATATACAATTGAGACTGGATGGATATATGATATGGAAAATGAAAGTGATAAATATTATTTCAA>6695
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
1019>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>1019
 390>TTAGAGTTGGGGAAGATGTATATTATTTTGGAGAAACATATACAATTGAGACTGGATGGATATATGATATGGAAAATGAAAGTGATAAATATTATTTCAA>489

         *         *         *         *         *         *         *         *         *         *
6696>TCCAGAAACTAAAAAAGCATGCAAAGGTATTAATTTAATTGATGATATAAAATATTATTTTGATGAGAAGGGCATAATGAGAACGGGTCTTATATCATTT>6795
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
1019>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>1019
 490>TCCAGAAACTAAAAAAGCATGCAAAGGTATTAATTTAATTGATGATATAAAATATTATTTTGATGAGAAGGGCATAATGAGAACGGGTCTTATATCATTT>589

         *         *         *         *         *         *         *         *         *         *
6796>GAAAATAATAATTATTACTTTAATGAGAATGGTGAAATGCAATTTGGTTATATAAATATAGAAGATAAGATGTTCTATTTTGGTGAAGATGGTGTCATGC>6895
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
1019>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>1019
 590>GAAAATAATAATTATTACTTTAATGAGAATGGTGAAATGCAATTTGGTTATATAAATATAGAAGATAAGATGTTCTATTTTGGTGAAGATGGTGTCATGC>689

         *         *         *         *         *         *         *         *         *         *
6896>AGATTGGAGTATTTAATACACCAGATGGATTTAAATACTTTGCACATCAAAATACTTTGGATGAGAATTTTGAGGGAGAATCAATAAACTATACTGGTTG>6995
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
1019>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>1019
 690>AGATTGGAGTATTTAATACACCAGATGGATTTAAATACTTTGCACATCAAAATACTTTGGATGAGAATTTTGAGGGAGAATCAATAAACTATACTGGTTG>789

         *         *         *          *         *         *         *         *         *         *
6996>GTTAGATTTAGATGAAAAGAGATATTATTTT-ACAGATGAATATATTGCAGCAACTGGTTCAGTTATTATTGATGGTGAGGAGTATTATTTTGATCCTGA>7094
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
```

```
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
1019>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>1019
 790>GTTAGATTTAGATGAAAAGAGATATTATTTTTACAGATGAATATATTGCAGCAACTGGTTCAGTTATTATTGATGGTGAGGAGTATTATTTTGATCCTGA>889

          *         *         *         *
7095>TACAGCTCAATTAGTGATTAGTGAAGggtacc~~~~~~~~~~>7126
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 949>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>949
 104<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<104
 147<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<147
 206<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<206
 879>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>879
  33<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<33
 512>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>512
  96<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<96
1019>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>1019
 890>TACAGCTCAATTAGTGATTAGTGAAGGGTACCGGCCGCATGC>931
```

APPENDIX II: rTcdB-GTD-NVN Sequence Alignment

```
rTcdB_GTD_NVN.ape (insert from pBMJ_70402; colony 6) from 295 to 2240
Alignment to
_B03_colony6__M13Reverse_015.seq--  Matches:792; Mismatches:0; Gaps:1154; Unattempted:0
_E02_colony6__B2F_004.seq--  Matches:271; Mismatches:0; Gaps:1675; Unattempted:0
_F02_colony6__B2R_003.seq--  Matches:808; Mismatches:0; Gaps:1138; Unattempted:0
_G02_colony6__B3F_002.seq--  Matches:76; Mismatches:0; Gaps:1870; Unattempted:0


        *         *         *         *         *         *         *         *         *         *
295>tgtacaATGAGTTTAGTTAATAGAAAACAGTTAGAAAAAATGGCAAATGTAAGATTTCGTACTCAAGAAGATGAATATGTTGCAATATTGGATGCTTTAG>394
 54>TGTACAATGAGTTTAGTTAATAGAAAACAGTTAGAAAAAATGGCAAATGTAAGATTTCGTACTCAAGAAGATGAATATGTTGCAATATTGGATGCTTTAG>153
647>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>647
842<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<842
 72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72


        *         *         *         *         *         *         *         *         *         *
395>AAGAATATCATAATATGTCAGAGAATACTGTAGTCGAAAAATATTTAAAATTAAAAGATATAAATAGTTTAACAGATATTTATATAGATACATATAAAAA>494
154>AAGAATATCATAATATGTCAGAGAATACTGTAGTCGAAAAATATTTAAAATTAAAAGATATAAATAGTTTAACAGATATTTATATAGATACATATAAAAA>253
647>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>647
842<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<842
 72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72


        *         *         *         *         *         *         *         *         *         *
495>ATCTGGTAGAAATAAAGCCTTAAAAAAATTTAAGGAATATCTAGTTACAGAAGTATTAGAGCTAAAGAATAATAATTTAACTCCAGTTGAGAAAAATTTA>594
254>ATCTGGTAGAAATAAAGCCTTAAAAAAATTTAAGGAATATCTAGTTACAGAAGTATTAGAGCTAAAGAATAATAATTTAACTCCAGTTGAGAAAAATTTA>353
647>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>647
842<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<842
 72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72


        *         *         *         *         *         *         *         *         *         *
595>CATTTTGTTTGGATTGGAGGTCAAATAAATGACACTGCTATTAATTATATAAATCAATGGAAAGATGTAAATAGTGATTATAATGTTAATGTTTTTTATG>694
354>CATTTTGTTTGGATTGGAGGTCAAATAAATGACACTGCTATTAATTATATAAATCAATGGAAAGATGTAAATAGTGATTATAATGTTAATGTTTTTTATG>453
647>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>647
842<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<842
 72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72


        *         *         *         *         *         *         *         *         *         *
695>ATAGTAATGCATTTTTGATAAACACATTGAAAAAAACTGTAGTAGAATCAGCAATAAATGATACACTTGAATCATTTAGAGAAAACTTAAATGACCCTAG>794
454>ATAGTAATGCATTTTTGATAAACACATTGAAAAAAACTGTAGTAGAATCAGCAATAAATGATACACTTGAATCATTTAGAGAAAACTTAAATGACCCTAG>553
647>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>647
842<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<842
 72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72


        *         *         *         *         *         *         *         *         *         *
795>ATTTGACTATAATAAATTCTTCAGAAAACGTATGGAAATAATTTATGATAAACAGAAAAATTTCATAAACTACTATAAAGCTCAAAGAGAAGAAAATCCT>894
554>ATTTGACTATAATAAATTCTTCAGAAAACGTATGGAAATAATTTATGATAAACAGAAAAATTTCATAAACTACTATAAAGCTCAAAGAGAAGAAAATCCT>653
647>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>647
842<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<842
 72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72


        *         *         *         *         *         *         *         *         *         *
895>GAACTTATAATTGATGATATTGTAAAGACATATCTTTCAAATGAGTATTCAAAGGAGATAGATGAACTTAATACCTATATTGAAGAATCCTTAAATAAAA>994
654>GAACTTATAATTGATGATATTGTAAAGACATATCTTTCAAATGAGTATTCAAAGGAGATAGATGAACTTAATACCTATATTGAAGAATCCTTAAATAAAA>753
647>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>647
842<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<842
 72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72


        *         *         *         *         *         *         *         *         *         *
995>TTACACAGAATAGTGGAAATGATGTTAGAAACTTTGAAGAATTTAAAAATGGAGAGTCATTCAACTTATATGAACAAGAGTTGGTAGAAAGGTGGAATTT>1094
754>TTACACAGAATAGTGGAAATGATGTTAGAAACTTTGAAGAATTTAAAAATGGAGAGTCATTCAACTTATATGAACAAGAGTTGGTAGAAAGG~~~~~~~~>845
648>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GTGGAATTT>656
842<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<842
 72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72
```

```
              *         *         *         *         *         *         *         *         *         *
1095>AGCTGCTGCTTCTGACATATTAAGAATATCTGCATTAAAAGAAATTGGTGGTATGTATTTAaATGTTaATATGTTACCAGGAATACAACCAGACTTATTT>1194
 845>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>845
 657>AGCTGCTGCTTCTGACATATTAAGAATATCTGCATTAAAAGAAATTGGTGGTATGTATTTAAATGTTAATATGTTACCAGGAATACAACCAGACTTATTT>756
 842<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<842
  72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72

              *         *         *         *         *         *         *         *         *         *
1195>GAGTCTATAGAGAAACCTAGTTCAGTAACAGTGGATTTTTGGGAAATGACAAAGTTAGAAGCTATAATGAAATACAAAGAATATATACCAGAATATACCT>1294
 845>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>845
 757>GAGTCTATAGAGAAACCTAGTTCAGTAACAGTGGATTTTTGGGAAATGACAAAGTTAGAAGCTATAATGAAATACAAAGAATATATACCAGAATATACCT>856
 842<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<842
  72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72

              *         *         *         *         *         *         *         *         *         *
1295>CAGAACATTTTGACATGTTAGACGAAGAAGTTCAAAGTAGTTTTGAATCTGTTCTAGCTTCTAAGTCAGATAAATCAGAAATATTCTCATCACTTGGTGA>1394
 845>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>845
 857>CAGAACATTTTGACATGTTAGACGAAGAAGTTCAAAGTAGTTTTGAATCTGTTCTAGCTTCT~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>918
 841<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~AAGTCAGATAAATCAGAAATATTCTCATCACTTGGTGA<804
  72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72

              *         *         *         *         *         *         *         *         *         *
1395>TATGGAGGCATCACCACTAGAAGTTAAAATTGCATTTAATAGTAAGGGTATTATAAATCAAGGGCTAATTTCTGTGAAAGACTCATATTGTAGCAATTTA>1494
 845>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>845
 918>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>918
 803<TATGGAGGCATCACCACTAGAAGTTAAAATTGCATTTAATAGTAAGGGTATTATAAATCAAGGGCTAATTTCTGTGAAAGACTCATATTGTAGCAATTTA<704
  72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72

              *         *         *         *         *         *         *         *         *         *
1495>ATAGTAAAACAAATCGAGAATAGATATAAAATATTGAATAATAGTTTAAATCCAGCTATTAGCGAGGATAATGATTTTAATACTACAACGAATACCTTTA>1594
 845>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>845
 918>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>918
 703<ATAGTAAAACAAATCGAGAATAGATATAAAATATTGAATAATAGTTTAAATCCAGCTATTAGCGAGGATAATGATTTTAATACTACAACGAATACCTTTA<604
  72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72

              *         *         *         *         *         *         *         *         *         *
1595>TTGATAGTATAATGGCTGAAGCTAATGCAGATAATGGTAGATTTATGATGGAACTAGGAAAGTATTTAAGAGTTGGTTTCTTCCCAGATGTTAAAACTAC>1694
 845>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>845
 918>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>918
 603<TTGATAGTATAATGGCTGAAGCTAATGCAGATAATGGTAGATTTATGATGGAACTAGGAAAGTATTTAAGAGTTGGTTTCTTCCCAGATGTTAAAACTAC<504
  72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72

              *         *         *         *         *         *         *         *         *         *
1695>TATTAACTTAAGTGGCCCTGAAGCATATGCGGCAGCTTATCAAGATTTATTAATGTTTAAAGAAGGCAGTATGAATATCCATTTGATAGAAGCTGATTTA>1794
 845>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>845
 918>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>918
 503<TATTAACTTAAGTGGCCCTGAAGCATATGCGGCAGCTTATCAAGATTTATTAATGTTTAAAGAAGGCAGTATGAATATCCATTTGATAGAAGCTGATTTA<404
  72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72

              *         *         *         *         *         *         *         *         *         *
1795>AGAAACTTTGAAATCTCTAAAACTAATATTTCTCAATCAACTGAACAAGAAATGGCTAGCTTATGGTCATTTGACGATGCAAGAGCTAAAGCTCAATTTG>1894
 845>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>845
 918>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>918
 403<AGAAACTTTGAAATCTCTAAAACTAATATTTCTCAATCAACTGAACAAGAAATGGCTAGCTTATGGTCATTTGACGATGCAAGAGCTAAAGCTCAATTTG<304
  72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72

              *         *         *         *         *         *         *         *         *         *
1895>AAGAATATAAAAGGAATTATTTTGAAGGTTCTCTTGGTGAAGATGATAATCTTGATTTTTCTCAAAATATAGTAGTTGACAAGGAGTATCTTTTAGAAAA>1994
 845>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>845
 918>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>918
 303<AAGAATATAAAAGGAATTATTTTGAAGGTTCTCTTGGTGAAGATGATAATCTTGATTTTTCTCAAAATATAGTAGTTGACAAGGAGTATCTTTTAGAAAA<204
  72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72
```

```
         *         *         *         *         *         *         *         *         *         *
1995>AATATCTTCATTAGCAAGAAGTTCAGAGAGAGGATATATACACTATATTGTTCAGTTACAAGGAGATAAAATTAGTTATGAAGCAGCATGTAACTTATTT>2094
 845>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>845
 918>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>918
 203<AATATCTTCATTAGCAAGAAGTTCAGAGAGAGGATATATACACTATATTGTTCAGTTACAAGGAGATAAAATTAGTTATGAAGCAGCATGTAACTTATTT<104
  72>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>72


         *         *         *         *         *         *         *         *         *         *
2095>GCAAAGACTCCTTATGaTAGTGTACTATTTCAGAAAAATATAGAAGATTCAGAAATTGCATATTATTATAATCCTGGAGATGGTGAAATACAAGAAATAG>2194
 845>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>845
 918>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>918
 103<GCAAAGACTCCTTATGATAGTGTACTATTTCAGAAAAATATAGAAGATTCAGAAATTGCATATTATTATA~~~~~~~~~~~~~~~~~~~~~~~~~~~~<34
  73>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ATCCTGGAGATGGTGAAATACAAGAAATAG>102


         *         *         *         *         *
2195>ACAAGTATAAAATTCCAAGTATAATTTCTGATAGACCTAAGATTAA>2240
 845>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>845
 918>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~>918
  34<~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~<34
 103>ACAAGTATAAAATTCCAAGTATAATTTCTGATAGACCTAAGATTAA>148
```

APPENDIX III: *pBMJ Plasmid Constructs*

Plasmid Name	Parent Vector	Insert	Size (bp)	Primers Used[c]	Restriction Enzymes Used	Resistance Marker
pBMJ_130401	pHis1522[a]	*tcdb*-WT[b]	14458	1 & 2	BsrG1/Kpn1	Amp (*E. coli*); Tet (*B. meg*)
pBMJ_70401	pCR-XL-TOPO	*tcdb*-GTD	5902	1 & 3	BsrG1/BsaB1	Kan
pBMJ_70402	pBMJ_70401 (pCR-XL-TOPO)	*tcdb*-GTD-NVN	5902	4 & 5 (SDM)	-	Kan

NOTE: Grey indicates protein expression plasmids. [a]i.e. pANK_80403: created by modifying pWH1520.[50]
[b]Isolated from *C. difficile* genomic DNA (ATTC 9689D-5 DR). [c]See **Table 2**.

REFERENCES

1. Goudarzi, M.; Seyedjavadi, S. S.; Goudarzi, H.; Aghdam, E. M.; Nazeri, S. *Clostridium difficile* Infection: Epidemiology, Pathogenesis, Risk Factors, and Therapeutic Options *Scientifica*. **2014,** *article ID 916826,* 9 pages.
2. Kelly, C. P.; LaMont, J. T. *Clostridium difficile* Infection *Annual Review of Medicine*. **1998,** *49,* 375-390.
3. Diaz-Gonzalez, F.; Milano, M.; Olguin-Araneda, V.; Pizarro-Cerda, J.; Castro-Cordova, P.; Tzeng, S.; Maier, C. S.; Sarker, M. R.; Paredes-Sabja, D. Protein Composition of the Outermost Exosporium-Like Layer of *Clostridium Difficile* 630 Spores *Journal of Proteomics*. **2015,** *123,* 1-13.
4. Wiggs, L. S. Centers for Disease Control and Prevention. Public Health Image Library. https://phil.cdc.gov/phil/details.asp?pid=9999
5. Carrico, R. M. Silent Menace. *C. difficile* and its Threat to Health Care *Health Facil Manage*. **2011,** *24,* 43-45.
6. Taylor, N. S.; Thorne, G. M.; Bartlett, J. G. Comparison of Two Toxins Produced by *Clostridium difficile Infect Immun*. **1981,** *34,* 1036-1043.
7. Kelly, C. P.; Pothoulakis, C.; LaMont, J. T. *Clostridium difficile* Colitis *The New England Journal of Medicine*. **1994,** *330,* 257-262.
8. Wilkins, T. D.; Lyerly, D. M. *Clostridium difficile* Toxins Attack Rho *Trends in Microbiology*. **1996,** *4,* 49-51.
9. Mergenhagen, K. A.; Wojciechowski, A. L.; Paladino, J. A. A Review of the Economics of Treating *Clostridium difficile* Infection *PharmacoEconomics*. **2014,** *32,* 639-650.
10. Rupnik, M.; Wilcox, M. H.; Gerding, D. N. *Clostridium difficile* infection: new developments

in epidemiology and pathogenesis *Nature Reviews: Microbiology*. **2009,** *7,* 526-536.

11. Abdeen, S. J. Development of Peptide Inhibitors Targeting *Clostridium difficile* Toxins A/B and Characterizing the Regulatory Role of a Putative Negative Regulator TcdC in *Clostridium difficile* Toxin Gene Expression. Ph.D. Thesis, Wayne State University, Detroit, MI, 2012.
12. Hung, Y.; Lee, J.; Lin, H.; Liu, H.; Wu, Y.; Tsai, P.; Ko, W. Clinical Impact of *Clostridium difficile* Colonization *Journal of Microbiology, Immunology and Infection*. **2015,** *48,* 241-248.
13. Centers for Disease Control and Prevention. CDC Newsroom: Nearly Half a Million Americans Suffered from *Clostridium difficile* Infections in a Single Year. http://www.cdc.gov/media/releases/2015/p0225-clostridium-difficile.html (accessed Nov 8, 2016).
14. Miller, B. A.; Chen, L. F.; Sexton, D. J.; Anderson, D. J. Comparison of the Burdens of Hospital-Onset, Healthcare Facility – Associated *Clostridium difficile* Infection due to Methicillin-Resistant Staphylococcus aureus in Community Hospitals *Infect Cont Hosp Ep*. **2011,** *32,* 387-390.
15. Jank, T.; Aktories, K. Structure and Mode of Action of Clostridial Glucosylating Toxins: The ABCD Model *Trends in Microbiology*. **2008,** *16,* 222-228.
16. Chumbler, N.; Rutherford, S.; Zhang, Z.; Farrow, M.; Lisher, J.; Farquhar, E.; Giedroc, D.; Spiller, B.; Melnyk, R.; Lacy, D. B. Crystal Structure of *Clostridium difficile* Toxin A *Nature Microbiology*. **2016,** *1,* 1-6.
17. Ho, J. G. S.; Greco, A.; Rupnik, M.; Ng, K. K. S. Crystal Structure of Receptor-Binding C-Terminal Repeats from *Clostridium difficile* Toxin A *Proc Natl Acad Sci USA*. **2005,** *102,* 18373-18378.
18. Greco, A.; Ho, J. G. S.; Lin, S.; Palcic, M. M.; Rupnik, M.; Ng, K. K. S. Carbohydrate

Recognition by *Clostridium difficile* Toxin A *Nature Structural & Molecular Biology*. **2006,** *13,* 460-461.

19. Martin-Verstraete, I.; Peltier, J.; Dupuy, B. The Regulatory Networks That Control *Clostridium difficile* Toxin Synthesis *Toxins*. **2016,** *8,* 153.
20. Carroll, K. C.; Bartlett, J. G. Biology of *Clostridium difficile*: Implications for Epidemiology and Diagnosis *Annual Review of Microbiology*. **2011,** *65,* 501-521.
21. Aktories, K. Self-Cutting to Kill: New Insights into the Processing of *Clostridium difficile* Toxins *ACS Chemical Biology*. **2007,** *2,* 228-230.
22. Cherian, R. M.; Jin, C.; Liu, J.; Karlsson, N. G.; Holgersson, J. Recombinant Mucin-Type Fusion Proteins with Galα1,3Gal Substitution as *Clostridium difficile* Toxin A Inhibitors *Infect Immun*. **2016,** *84,* 2842-2852.
23. Voth, D. E.; Ballard, J. D. *Clostridium difficile* Toxins: Mechanism of Action and Role in Disease *Clinical Microbiology Reviews*. **2005,** *18,* 247-263.
24. Ivarsson, M. E.; Leroux, J.; Castagner, B. Targeting Bacterial Toxins *Angew. Chem. Int. Ed.* **2012,** *51,* 4024-4045.
25. Li, S.; Shi, L.; Yang, Z.; Zhang, Y.; Perez-Cordon, G.; Huang, T.; Ramsey, J.; Oezguen, N.; Savidge, T. C.; Feng, H. Critical Roles of *Clostridium difficile* Toxin B Enzymatic Activities in Pathogenesis *Infect Immun*. **2015,** *83,* 502-513.
26. Melnyk, R. A.; Collier, R. J. A Loop Network within the Anthrax Toxin Pore Positions the Phenylalanine Clamp in an Active Conformation *Proc Natl Acad Sci USA*. **2006,** *103,* 9802-9807.
27. Just, I.; Selzer, J.; Wilm, M.; von Eichel-Streiber, C.; Mann, M.; Aktories, K. Glucosylation of Rho Proteins by *Clostridium difficile* Toxin B *Nature*. **1995,** *375,* 500-503.

28. Bibbo, S.; Lopetuso, L. R.; Ianiro, G.; Di Rienzo, T.; Gasbarrini, A.; Cammarota, G. Role of Microbiota and Innate Immunity in Recurrent *Clostridium difficile* Infection *Journal of Immunology Research*. **2014,** *article ID 462740,* 8 pages.

29. Jafari, N. V.; Kuehne, S. A.; Bryant, C. E.; Elawad, M.; Wren, B. W.; Minton, N. P.; Allan, E.; Bajaj-Elliott, M. *Clostridium difficile* Modulates Host Innate Immunity via Toxin-Independent and Dependent Mechanism(s) *PLoS ONE*. **2013,** *8,* e69846.

30. Horst Ibelgaufts' COPE: Cytokines & Cells Online Pathfinder Encyclopedia. Cytokines. http://www.copewithcytokines.de/cope.cgi?key=cytokines (accessed Mar 28, 2015).

31. Opal, S. M.; DePalo, V. A. Anti-Inflammatory Cytokines *CHEST*. **2000,** *117,* 1162-1172.

32. El Feghaly, R. E.; Stauber, J. L.; Deych, E.; Gonzalez, C.; Tarr, P. I.; Haslam, D. B. Marker of Intestinal Inflammation, Not Bacterial Burden, Correlate With Clinical Outcomes in *Clostridium difficile* Infection *Clinical Infectious Diseases*. **2013,** *56,* 1713-1721.

33. Aslam, S.; Hamill, R. J.; Musher, D. M. Treatment of *Clostridium difficile*-Associated Disease: Old Therapies and New Strategies *Lancet Infect Dis*. **2005,** *5,* 549-557.

34. Leffler, D. A.; Lamont, J. T. Treatment of *Clostridium difficile*-Associated Diseases *Gastroenterology*. **2009,** *136,* 1899-1912.

35. Alang, N.; Kelly, C. Weight Gain After Fecal Microbiota Transplantation *Open Forum Infect Diseas*. **2014,** 1-2.

36. Evrensel, A.; Ceylan, M. E. The Gut-Brain Axis: The Missing Link to Depression *Clinical Psychopharmacology and Neuroscience*. **2015,** *13,* 239-244.

37. Cowardin, C. A.; Kuehne, S. A.; Buonomo, E. L.; Marie, C. S.; Minton, N. P.; Petri, W. A. Inflammasome Activation Contributes to Interleukin-23 Production in Response to *Clostridium difficile American Society for Microbiology*. **2015,** *6,* e02386-14.

38. Cowardin, C. A.; Jackman, B. M.; Noor, Z.; Burgess, S. L.; Feig, A. L.; Petri, W. A. Glucosylation Drives the Innate Inflammatory Response to *Clostridium difficile* Toxin A *Infect Immun*. **2016,** *84,* 2317-2323.

39. Chen, S.; Wang, H.; Gu, H.; Sun, C.; Li, S.; Feng, H.; Wang, J. Identification of an Essential Region for Translocation of *Clostridium difficile* Toxin B *Toxins*. **2016,** *8,* 241.

40. Schnell, L.; Mittler, A.; Mattarei, A.; Tehran, D. A.; Montecucco, C.; Barth, H. Semicarbazone EGA Inhibits Uptake of Diptheria Toxin into Human Cells and Protects Cells from Intoxication *Toxins*. **2016,** *8,* 221.

41. Tam, J.; Beilhartz, G. L.; Auger, A.; Gupta, P.; Therien, A. G.; Melnyk, R. A. Small Molecule Inhibitors of *Clostridium difficile* Toxin B-Induced Cellular Damage *Chem Biol*. **2015,** *22,* 175-185.

42. Puri, A. W.; Lupardus, P. J.; Deu, E.; Albow, V. E.; Garcia, K. C.; Bogyo, M. Rational Design of Inhibitors and Activity-Based Probes Targeting *Clostridium difficile* Virulence Factor TcdB *Chem Biol*. **2010,** *17,* 1201-1211.

43. Egerer, M.; Giesemann, T.; Jank, T.; Fullner Satchell, K. J.; Aktories, K. Auto-catalytic Cleavage of *Clostridium difficile* Toxins A and B Depends on Cysteine Protease Activity *J Biol Chem*. **2007,** *282,* 25314-25321.

44. Reineke, J.; Tenzer, S.; Rupnik, M.; Koschinski, A.; Hasselmayer, O.; Schrattenholz, A.; Schild, H.; von Eichel-Streiber, C. Autocatalytic Cleavage of *Clostridium difficile* Toxin B *Nature*. **2007,** *446,* 415-419.

45. Saul, R.; Ghidoni, J. J.; Molyneux, R. J.; Elbein, A. D. Castanospermine Inhibits Alpha-Glucosidase Activites and Alters Glycogen Distribution in Animals *Proc Natl Acad Sci USA*. **1985,** *82,* 93-97.

46. Abdeen, S. J.; Swett, R. J.; Feig, A. L. Peptide Inhibitors Targeting *Clostridium difficile* Toxins A and B *ACS Chemical Biology*. **2010,** *5,* 1097-1103.

47. Swett, R. J.; Cisneros, G. A.; Feig, A. L. Disruption of Intrinsic Motions as a Mechanism for Enzyme Inhibition *Biophysical Journal*. **2013,** *105,* 494-501.

48. Olling, A.; Hüls, C.; Goy, S.; Müller, M.; Krooss, S.; Rudolf, I.; Tatge, H.; Gerhard, R. The Combined Repetitive Oligopeptides of *Clostridium difficile* Toxin A Counteract Premature Cleavage of the Glucosyl-Transferase Domain by Stabilizing Protein Conformation *Toxins*. **2014,** *6,* 2162-2176.

49. Pruitt, R. N.; Chambers, M. G.; Ng, K. K.; Ohi, M. D.; Lacy, D. B. Structural Organization of the Functional Domains of *Clostridium difficile* Toxins A and B *Proc Natl Acad Sci USA*. **2010,** *107,* 13467–13472.

50. Kerzmann, A. N. Mechanistic Analysis of *Clostridium difficile* Toxin A. Ph.D. Thesis, Wayne State University, Detroit, MI, 2009.

51. Pruitt, R. N.; Chumbler, N. M.; Rutherford, S. A.; Farrow, M. A.; Friedman, D. B.; Spiller, B.; Lacy, D. B. *Clostridium difficile* Toxin A (TcdA) Glucosyltransferase Domain Bound to UDP-Glucose *J Biol Chem*. **2012,** *287,* 8013-8020.

52. Krippendorff, B.; Neuhaus, R.; Lienau, P.; Reichel, A.; Huisinga, W. Mechanism-Based Inhibition: Deriving K_I and k_{inact} Directly from Time-Dependent IC_{50} Values *Journal of Biomolecular Screening*. **2009,** *14,* 913-923.

53. Buonomo, E. L.; Madan, R.; Pramoonjago, P.; Li, L.; Okusa, M. D.; Petri, W. A. Role of Interleukin 23 Signaling in *Clostridium difficile* Colitis *Journal of Infectious Diseases*. **2013,** *208,* 917-920.

54. Busch, C.; Hofman, F.; Selzer, J.; Munro, S.; Jeckel, D.; Aktories, K. Enzyme Activity of

Large A Common Motif of Eukaryotic Glycosyltransferases is Essential for the Enzyme Activity of Clostridial Cytotoxins *J Biol Chem*. **1998,** *273,* 19566-19572.

55. Borish, L.; Rosenwasser, L. J. Update on Cytokines *J Allergy Clin Immunol*. **1996,** *97,* 719-734.

56. Denes, A.; Lopez-Castejon, G.; Brough, D. Caspase-1: is IL-1 Just the Tip of the ICEberg? *Cell Death and Disease*. **2012,** *3,* e338.

57. Hannelien, V.; Karel, G.; Jo, V. D.; Sofie, S. The Role of CXC Chemokines in the Transition of Chronic Inflammation to Esophageal and Gastric Cancer *Biochimica et Biophysica Acta*. **2012,** *1825,* 117-129.

58. Lawrence, T. The Nuclear Factor NF-ϰB Pathway in Inflammation *Cold Spring Harb Perspect Biol*. **2009,** *1,* a001651.

59. Gilmore, T. D. Introduction to NF-ϰB: Players, Pathways, Perspectives *Oncogene*. **2006,** *25,* 6680-6684.

60. Arduengo, M. CellTiter-Glo™ Luminescent Cell Viability Assay for Cytotoxicity and Cell Proliferation Studies *Cell Notes*. **2003,** *5,* 5-17.

61. InvivoGen: Innovation within Reach. RAW-Blue™ Cells. http://www.invivogen.com/raw-blue (accessed June 6, 2015).

62. InvivoGen: Innovation within Reach. QUANTI-Blue™. http://www.invivogen.com/quanti-blue (accessed June 6, 2015).

63. Tseng, W.; Leong, X.; Engleman, E. Orthotopic Mouse Model of Colorectal Cancer *Journal of Visualized Experiments*. **2007,** *10,* 484.

64. Zhang, Z.; Park, M.; Tam, J.; Auger, A.; Beilhartz, G. L.; Lacy, D. B.; Melnyk, R. A. Translocation Domain Mutations Affecting Cellular Toxicity Identify the *Clostridium difficile*

Toxin B Pore *Proc Natl Acad Sci USA*. **2014,** *111,* 3721-3726.

65. Chumbler, N. M.; Farrow, M. A.; Lapierre, L. A.; Franklin, J. L.; Haslam, D. B.; Goldenring, J. R.; Lacy, D. B. *Clostridium difficile* Toxin B Cause Epithelial Cell Necrosis Through an Autoprocessing-Independent Mechanism *PLoS Pathog*. **2012,** *8,* e1003072.
66. Farrow, M. A.; Chumbler, N. M.; Lapierre, L. A.; Franklin, J. L.; Rutherford, S. A.; Goldenring, J. R.; Lacy, D. B. *Clostridium difficile* Toxin B-Induced Necrosis is Mediated by the Host Epithelial Cell NADPH Oxidase Complex *Proc Natl Acad Sci USA*. **2013,** *110,* 18674-18679.
67. Ryder, A. B.; Huang, Y.; Li, H.; Zheng, M.; Wang, X.; Stratton, C. W.; Xu, X.; Tang, Y. Assessment of *Clostridium difficile* Infections by Quantitative Detection of tcdB Toxin by Use of a Real-Time Cell Analysis System *Journal of Clinical Microbiology*. **2010,** *48,* 4129-4134.
68. Stabler, R. A.; Dawson, L. F.; Phua, L. T. H.; Wren, B. W. Comparative Analysis of BI/NAP1/027 Hypervirulent Strains Reveals Novel Toxin B-Encoding Gene (tcdB) Sequences *Journal of Medical Microbiology*. **2008,** *57,* 771-775.
69. Warny, M.; Pepin, J.; Fang, A.; Killgore, G.; Thompson, A.; Brazier, J.; Frost, E.; McDonald, L. C. Toxin Production of an Emerging Strain of *Clostridium difficile*-Associated with Outbreaks of Severe Disease in North America and Europe *Lancet*. **2005,** *366,* 1079-1084.
70. New England Biolabs, Inc. Q5® High-Fidelity DNA Polymerase. https://www.neb.com/products/m0491-q5-high-fidelity-dna-polymerase (accessed Nov 13, 2016).
71. New England Biolabs, Inc. PCR Selection Tool. https://www.neb.com/tools-and-resources/selection-charts/pcr-selection-chart (accessed Nov 13, 2016).
72. Burger, S.; Tatge, H.; Hofmann, F.; Genth, H.; Just, I.; Gerhard, R. Expression of Recombinant

Clostridium difficile Toxin A using the *Bacillus megaterium* System *Biochem Biophys Res Commun*. **2003,** *307,* 584-588.

73. Yang, G.; Zhou, B.; Wang, J.; He, X.; Sun, X.; Nie, W.; Tzipori, S.; Feng, H. Expression of Recombinant *Clostridium difficile* Toxin A and B in *Bacillus megaterium BMC Microbiology*. **2008,** *8,* 192.

ABSTRACT

Clostridium difficile (*C. diff*) is an anaerobic, spore-forming pathogen responsible for hospital-acquired *C. diff*-associated diseases (CDAD) that are extremely common following patient treatment with broad-spectrum antibiotics. Symptoms of CDAD range from excessive diarrhea, colonic inflammation, pseudomembranous colitis, and in extreme cases, death. Although enterotoxins A and B (TcdA and TcdB, respectively) are the main virulence factors involved and antibiotic-resistant strains are continually emerging today, antibiotic treatment remains the most common method against CDAD. The work presented here details a potential anti-toxin therapeutic utilizing an epoxide-containing peptide (NH_2-HQSPG_{epoxy}HHGGGC-$CONH_2$) that covalently crosslinks the enzymatic domain of TcdA and has shown to inhibit glucosyltransferase activity in a concentration and time-dependent manner. While further analysis is still needed, anti-toxin, peptide-based agents as such show promising potential for an advanced remedy against CDAD.